PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

A9

(51) International Patent Classification 6 :	A2	(11) International Publication Number: WO 99/31236
C12N 15/12, C07K 14/47, 16/18, C12Q 1/68		(43) International Publication Date: 24 June 1999 (24.06.99)

(21) International Application Number: PCT/IB98/02122

(22) International Filing Date: 17 December 1998 (17.12.98)

(30) Priority Data:
60/069,957	17 December 1997 (17.12.97)	US
60/074,121	9 February 1998 (09.02.98)	US
60/081,563	13 April 1998 (13.04.98)	US
60/096,116	10 August 1998 (10.08.98)	US

(71) Applicant (for all designated States except US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR).

(72) Inventors; and
(75) Inventors/Applicants (for US only): BOUGUELERET, Lydie [FR/FR]; 108, avenue Victor Hugo, F-92170 Vanves (FR). DUCLERT, Aymeric [FR/FR]; 6 ter, rue Victorine, F-94100 Saint-Maur (FR). DUMAS MILNE EDWARDS, Jean-Baptiste [FR/FR]; 8, rue Grégoire de Tours, F-75006 Paris (FR).

(74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: EXTENDED cDNAs FOR SECRETED PROTEINS

(57) Abstract

The sequences of extended cDNAs encoding secreted proteins are disclosed. The extended cDNAs can be used to express secreted proteins or portions thereof or to obtain antibodies capable of specifically binding to the secreted proteins. The extended cDNAs may also be used in diagnostic, forensic, gene therapy, and chromosome mapping procedures. The extended cDNAs may also be used to design expression vectors and secretion vectors.

EXTENDED cDNAS for secreted proteins

The present application relates to extended cDNAs which were disclosed in several United States Provisional Patent Applications. Table I lists the SEQ ID Nos. of the extended cDNAs in the present application, the SEQ ID Nos. of the identical or nearly identical extended cDNAs in the provisional applications, and the identities of the provisional
5 applications in which the extended cDNAs were disclosed.

Background of the Invention

The estimated 50,000-100,000 genes scattered along the human chromosomes offer tremendous promise for the understanding, diagnosis, and treatment of human diseases. In addition, probes capable of specifically hybridizing to loci distributed throughout the human genome find applications in the construction of high resolution chromosome maps
10 and in the identification of individuals.

In the past, the characterization of even a single human gene was a painstaking process, requiring years of effort. Recent developments in the areas of cloning vectors, DNA sequencing, and computer technology have merged to greatly accelerate the rate at which human genes can be isolated, sequenced, mapped, and characterized. Cloning vectors such as yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) are able to accept DNA
15 inserts ranging from 300 to 1000 kilobases (kb) or 100-400 kb in length respectively, thereby facilitating the manipulation and ordering of DNA sequences distributed over great distances on the human chromosomes. Automated DNA sequencing machines permit the rapid sequencing of human genes. Bioinformatics software enables the comparison of nucleic acid and protein sequences, thereby assisting in the characterization of human gene products.

Currently, two different approaches are being pursued for identifying and characterizing the genes distributed
20 along the human genome. In one approach, large fragments of genomic DNA are isolated, cloned, and sequenced. Potential open reading frames in these genomic sequences are identified using bio-informatics software. However, this approach entails sequencing large stretches of human DNA which do not encode proteins in order to find the protein encoding sequences scattered throughout the genome. In addition to requiring extensive sequencing, the bio-informatics software may mischaracterize the genomic sequences obtained. Thus, the software may produce false positives in which
25 which non-coding DNA is mischaracterized as coding DNA or false negatives in which coding DNA is mislabeled as non-coding DNA.

An alternative approach takes a more direct route to identifying and characterizing human genes. In this approach, complementary DNAs (cDNAs) are synthesized from isolated messenger RNAs (mRNAs) which encode human proteins. Using this approach, sequencing is only performed on DNA which is derived from protein coding portions of the
30 genome. Often, only short stretches of the cDNAs are sequenced to obtain sequences called expressed sequence tags (ESTs). The ESTs may then be used to isolate or purify extended cDNAs which include sequences adjacent to the EST sequences. The extended cDNAs may contain all of the sequence of the EST which was used to obtain them or only a portion of the sequence of the EST which was used to obtain them. In addition, the extended cDNAs may contain the full coding sequence of the gene from which the EST was derived or, alternatively, the extended cDNAs may include

portions of the coding sequence of the gene from which the EST was derived. It will be appreciated that there may be several extended cDNAs which include the EST sequence as a result of alternate splicing or the activity of alternative promoters.

In the past, the short EST sequences which could be used to isolate or purify extended cDNAs were often

5   obtained from oligo-dT primed cDNA libraries. Accordingly, they mainly corresponded to the 3' untranslated region of the mRNA. In part, the prevalence of EST sequences derived from the 3' end of the mRNA is a result of the fact that typical techniques for obtaining cDNAs, are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs. (Adams et al., *Nature* 377:174, 1996, Hillier et al., *Genome Res.* 6:807-828, 1996).

In addition, in those reported instances where longer cDNA sequences have been obtained, the reported

10   sequences typically correspond to coding sequences and do not include the full 5' untranslated region of the mRNA from which the cDNA is derived. Such incomplete sequences may not include the first exon of the mRNA, particularly in situations where the first exon is short. Furthermore, they may not include some exons, often short ones, which are located upstream of splicing sites. Thus, there is a need to obtain sequences derived from the 5' ends of mRNAs which can be used to obtain extended cDNAs which may include the 5' sequences contained in the 5' ESTs.

15   While many sequences derived from human chromosomes have practical applications, approaches based on the identification and characterization of those chromosomal sequences which encode a protein product are particularly relevant to diagnostic and therapeutic uses. Of the 50,000-100,000 protein coding genes, those genes encoding proteins which are secreted from the cell in which they are synthesized, as well as the secreted proteins themselves, are particularly valuable as potential therapeutic agents. Such proteins are often involved in cell to cell communication and

20   may be responsible for producing a clinically relevant response in their target cells.

In fact, several secretory proteins, including tissue plasminogen activator, G-CSF, GM-CSF, erythropoietin, human growth hormone, insulin, interferon-α, interferon-β, interferon-γ, and interleukin-2, are currently in clinical use. These proteins are used to treat a wide range of conditions, including acute myocardial infarction, acute ischemic stroke, anemia, diabetes, growth hormone deficiency, hepatitis, kidney carcinoma, chemotherapy induced neutropenia and

25   multiple sclerosis. For these reasons, extended cDNAs encoding secreted proteins or portions thereof represent a particularly valuable source of therapeutic agents. Thus, there is a need for the identification and characterization of secreted proteins and the nucleic acids encoding them.

In addition to being therapeutically useful themselves, secretory proteins include short peptides, called signal peptides, at their amino termini which direct their secretion. These signal peptides are encoded by the signal sequences

30   located at the 5' ends of the coding sequences of genes encoding secreted proteins. Because these signal peptides will direct the extracellular secretion of any protein to which they are operably linked, the signal sequences may be exploited to direct the efficient secretion of any protein by operably linking the signal sequences to a gene encoding the protein for which secretion is desired. This may prove beneficial in gene therapy strategies in which it is desired to deliver a particular gene product to cells other than the cell in which it is produced. Signal sequences encoding signal peptides

also find application in simplifying protein purification techniques. In such applications, the extracellular secretion of the desired protein greatly facilitates purification by reducing the number of undesired proteins from which the desired protein must be selected. Thus, there exists a need to identify and characterize the 5′ portions of the genes for secretory proteins which encode signal peptides.

5   Public information on the number of human genes for which the promoters and upstream regulatory regions have been identified and characterized is quite limited. In part, this may be due to the difficulty of isolating such regulatory sequences. Upstream regulatory sequences such as transcription factor binding sites are typically too short to be utilized as probes for isolating promoters from human genomic libraries. Recently, some approaches have been developed to isolate human promoters. One of them consists of making a CpG island library (Cross, S.H. et al.,

10   Purification of CpG Islands using a Methylated DNA Binding Column, Nature Genetics 6: 236-244 (1994)). The second consists of isolating human genomic DNA sequences containing SpeI binding sites by the use of SpeI binding protein. (Mortlock et al., *Genome Res.* 6:327-335, 1996). Both of these approaches have their limits due to a lack of specificity or of comprehensiveness.

        5′ ESTs and extended cDNAs obtainable therefrom may be used to efficiently identify and isolate upstream

15   regulatory regions which control the location, developmental stage, rate, and quantity of protein synthesis, as well as the stability of the mRNA. (Theil *et al.; BioFactors* 4:87-93, (1993). Once identified and characterized, these regulatory regions may be utilized in gene therapy or protein purification schemes to obtain the desired amount and locations of protein synthesis or to inhibit, reduce, or prevent the synthesis of undesirable gene products.

        In addition, ESTs containing the 5′ ends of secretory protein genes or extended cDNAs which include

20   sequences adjacent to the sequences of the ESTs may include sequences useful as probes for chromosome mapping and the identification of individuals. Thus, there is a need to identify and characterize the sequences upstream of the 5′ coding sequences of genes encoding secretory proteins.

Summary of the Invention

        The present invention relates to purified, isolated, or recombinant extended cDNAs which encode secreted

25   proteins or fragments thereof. Preferably, the purified, isolated or recombinant cDNAs contain the entire open reading frame of their corresponding mRNAs, including a start codon and a stop codon. For example, the extended cDNAs may include nucleic acids encoding the signal peptide as well as the mature protein. Alternatively, the extended cDNAs may contain a fragment of the open reading frame. In some embodiments, the fragment may encode only the sequence of the mature protein. Alternatively, the fragment may encode only a portion of the mature protein. A further aspect of the

30   present invention is a nucleic acid which encodes the signal peptide of a secreted protein.

        The present extended cDNAs were obtained using ESTs which include sequences derived from the authentic 5′ ends of their corresponding mRNAs. As used herein the terms "EST" or "5′ EST" refer to the short cDNAs which were used to obtain the extended cDNAs of the present invention. As used herein, the term "extended cDNA" refers to the cDNAs which include sequences adjacent to the 5′ EST used to obtain them. The extended cDNAs may contain all or a

-4-

portion of the sequence of the EST which was used to obtain them. The term "corresponding mRNA" refers to the mRNA which was the template for the cDNA synthesis which produced the 5' EST. As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual extended cDNA clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from

5 these clones could not be obtained directly either from the library or from total human DNA. The extended cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an

10 approximately 10^4-10^6 fold purification of the native message. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

As used herein, the term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living

15 animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated.

As used herein, the term "recombinant" means that the extended cDNA is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the extended cDNAs will represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone

20 molecules according to the present invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Preferably, the enriched extended cDNAs represent 15% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. More preferably, the enriched extended cDNAs represent 50% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. In a highly preferred

25 embodiment, the enriched extended cDNAs represent 90% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. "Stringent", "moderate," and "low" hybridization conditions are as defined in Example 29.

Unless otherwise indicated, a "complementary" sequence is fully complementary. Thus, extended cDNAs encoding secreted polypeptides or fragments thereof which are present in cDNA libraries in which one or more extended

30 cDNAs encoding secreted polypeptides or fragments thereof make up 5% or more of the number of nucleic acid inserts in the backbone molecules are "enriched recombinant extended cDNAs" as defined herein. Likewise, extended cDNAs encoding secreted polypeptides or fragments thereof which are in a population of plasmids in which one or more extended cDNAs of the present invention have been inserted such that they represent 5% or more of the number of inserts in the plasmid backbone are " enriched recombinant extended cDNAs" as defined herein. However, extended

cDNAs encoding secreted polypeptides or fragments thereof which are in cDNA libraries in which the extended cDNAs encoding secreted polypeptides or fragments thereof constitute less than 5% of the number of nucleic acid inserts in the population of backbone molecules, such as libraries in which backbone molecules having a cDNA insert encoding a secreted polypeptide are extremely rare, are not "enriched recombinant extended cDNAs."

5   In particular, the present invention relates to extended cDNAs which were derived from genes encoding secreted proteins. As used herein, a "secreted" protein is one which, when expressed in a suitable host cell, is transported across or through a membrane, including transport as a result of signal peptides in its amino acid sequence. "Secreted" proteins include without limitation proteins secreted wholly (e.g. soluble proteins), or partially (e.g. receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins which are
10   transported across the membrane of the endoplasmic reticulum.

Extended cDNAs encoding secreted proteins may include nucleic acid sequences, called signal sequences, which encode signal peptides which direct the extracellular secretion of the proteins encoded by the extended cDNAs. Generally, the signal peptides are located at the amino termini of secreted proteins.

Secreted proteins are translated by ribosomes associated with the "rough" endoplasmic reticulum. Generally,
15   secreted proteins are co-translationally transferred to the membrane of the endoplasmic reticulum. Association of the ribosome with the endoplasmic reticulum during translation of secreted proteins is mediated by the signal peptide. The signal peptide is typically cleaved following its co-translational entry into the endoplasmic reticulum. After delivery to the endoplasmic reticulum, secreted proteins may proceed through the Golgi apparatus. In the Golgi apparatus, the proteins may undergo post-translational modification before entering secretory vesicles which transport them across the
20   cell membrane.

The extended cDNAs of the present invention have several important applications. For example, they may be used to express the entire secreted protein which they encode. Alternatively, they may be used to express portions of the secreted protein. The portions may comprise the signal peptides encoded by the extended cDNAs or the mature proteins encoded by the extended cDNAs (i.e. the proteins generated when the signal peptide is cleaved off). The
25   portions may also comprise polypeptides having at least 10 consecutive amino acids encoded by the extended cDNAs. Alternatively, the portions may comprise at least 15 consecutive amino acids encoded by the extended cDNAs. In some embodiments, the portions may comprise at least 25 consecutive amino acids encoded by the extended cDNAs. In other embodiments, the portions may comprise at least 40 amino acids encoded by the extended cDNAs.

Antibodies which specifically recognize the entire secreted proteins encoded by the extended cDNAs or
30   fragments thereof having at least 10 consecutive amino acids, at least 15 consecutive amino acids, at least 25 consecutive amino acids, or at least 40 consecutive amino acids may also be obtained as described below. Antibodies which specifically recognize the mature protein generated when the signal peptide is cleaved may also be obtained as described below. Similarly, antibodies which specifically recognize the signal peptides encoded by the extended cDNAs may also be obtained.

·6·

In some embodiments, the extended cDNAs include the signal sequence. In other embodiments, the extended cDNAs may include the full coding sequence for the mature protein (i.e. the protein generated when the signal polypeptide is cleaved off). In addition, the extended cDNAs may include regulatory regions upstream of the translation start site or downstream of the stop codon which control the amount, location, or developmental stage of gene

5   expression. As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the cDNAs may be useful in treating or controlling a variety of human conditions. The extended cDNAs may also be used to obtain the corresponding genomic DNA. The term "corresponding genomic DNA" refers to the genomic DNA which encodes mRNA which includes the sequence of one of the strands of the extended cDNA in which thymidine residues in the sequence of the extended cDNA are replaced by uracil residues in the mRNA.

10   The extended cDNAs or genomic DNAs obtained therefrom may be used in forensic procedures to identify individuals or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal expression of the genes corresponding to the extended cDNAs. In addition, the present invention is useful for constructing a high resolution map of the human chromosomes.

The present invention also relates to secretion vectors capable of directing the secretion of a protein of

15   interest. Such vectors may be used in gene therapy strategies in which it is desired to produce a gene product in·one cell which is to be delivered to another location in the body. Secretion vectors may also facilitate the purification of desired proteins.

The present invention also relates to expression vectors capable of directing the expression of an inserted gene in a desired spatial or temporal manner or at a desired level. Such vectors may include sequences upstream of the

20   extended cDNAs such as promoters or upstream regulatory sequences.

In addition, the present invention may also be used for gene therapy to control or treat genetic diseases. Signal peptides may also be fused to heterologous proteins to direct their extracellular secretion.

One embodiment of the present invention is a purified or isolated nucleic acid comprising the sequence of one of SEQ ID NOs: 40-140 and 242-377 or a sequence complementary thereto. In one aspect of this embodiment, the

25   nucleic acid is recombinant.

Another embodiment of the present invention is a purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 40-140 and 242-377 or one of the sequences complementary thereto. In one aspect of this embodiment, the nucleic acid comprises at least 15, 25, 30, 40, 50, 75, or 100 consecutive bases of one of the sequences of SEQ ID NOs: 40-140 and 242-377 or one of the sequences complementary

30   thereto. The nucleic acid may be a recombinant nucleic acid.

Another embodiment of the present invention is a purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 40-140 and 242-377 or a sequence complementary to one of the sequences of SEQ ID NOs: 40-140 and 242-377. In one aspect of this embodiment, the nucleic acid is recombinant. '

-7-

Another embodiment of the present invention is a purified or isolated nucleic acid comprising the full coding sequences of one of SEQ ID NOs: 40-140 and 242-377, wherein the full coding sequence optionally comprises the sequence encoding signal peptide as well as the sequence encoding mature protein. In a preferred embodiment, the isolated or purified nucleic acid comprises the full coding sequence of one of SEQ ID Nos. 40, 42-44, 46, 48, 49, 51, 53,

5   60, 62-72, 76-78, 80-83, 85-88, 90, 93, 94, 97, 99-102, 104, 107-125, 127, 132, 135-138, 140 and 242-377 wherein the full coding sequence comprises the sequence encoding signal peptide and the sequence encoding mature protein. In one aspect of this embodiment, the nucleic acid is recombinant.

A further embodiment of the present invention is a purified or isolated nucleic acid comprising the nucleotides of one of SEQ ID NOs: 40-140 and 242-377 which encode a mature protein. In a preferred embodiment, the purified or

10   isolated nucleic acid comprises the nucleotides of one of SEQ ID NOs: 40-44, 46, 48, 49, 51-53, 55, 56, 58-72, 75-78, 80-88, 90, 93, 94, 97, 99-125, 127, 132, 133, 135-138, 140, and 242-377 which encode a mature protein. In one aspect of this embodiment, the nucleic acid is recombinant.

Yet another embodiment of the present invention is a purified or isolated nucleic acid comprising the nucleotides of one of SEQ ID NOs: 40-140 and 242-377 which encode the signal peptide. In a preferred embodiment,

15   the purified or isolated nucleic acid comprises the nucleotides of SEQ ID NOs: 40, 42-46, 48, 49, 51, 53, 57, 60, 62-73, 76-78, 80-83, 85-88, 90, 93-95, 97, 99-102, 104, 107-125, 127, 128, 130, 132, 134-140 and 242-377 which encode the signal peptide. In one aspect of this embodiment, the nucleic acid is recombinant.

Another embodiment of the present invention is a purified or isolated nucleic acid encoding a polypeptide having the sequence of one of the sequences of SEQ ID NOs: 141-241 and 378-513.

20   Another embodiment of the present invention is a purified or isolated nucleic acid encoding a polypeptide having the sequence of a mature protein included in one of the sequences of SEQ ID NOs: 141-241 and 378-513. In a preferred embodiment, the purified or isolated nucleic acid encodes a polypeptide having the sequence of a mature protein included in one of the sequences of SEQ ID NOs: 141-145, 147, 149, 150, 152-154, 156, 157, 159-172, 176-179, 181-189, 191, 194, 195, 198, 200-226, 228, 233, 234, 236-239, 241 and 378-513.

25   Another embodiment of the present invention is a purified or isolated nucleic acid encoding a polypeptide having the sequence of a signal peptide included in one of the sequences of SEQ ID NOs: 141-241 and 378-513. In a preferred embodiment, the purified or isolated nucleic acid encodes a polypeptide having the sequence of a signal peptide included in one of the sequences of SEQ ID NOs: 141, 143-147, 149, 150, 152, 154, 158, 161, 163-174, 177-179, 181-184, 186-189, 191, 194-196, 198, 200-203, 205, 208-226, 228, 229, 231, 233, 235-241, and 378-513.

30   Yet another embodiment of the present invention is a purified or isolated protein comprising the sequence of one of SEQ ID NOs: 141-241 and 378-513.

Another embodiment of the present invention is a purified or isolated polypeptide comprising at least 10 consecutive amino acids of one of the sequences of SEQ ID NOs: 141-241 and 378-513. In one aspect of this embodiment, the purified or isolated polypeptide comprises at least 15, 20, 25, 35, 50, 75, 100, 150 or 200 consecutive

-8-

amino acids of one of the sequences of SEQ ID NOs: 141-241 and 378-513. In still another aspect, the purified or

isolated polypeptide comprises at least 25 consecutive amino acids of one of the sequences of SEQ ID NOs: 141-241

and 378-513.

Another embodiment of the present invention is an isolated or purified polypeptide comprising a signal peptide

5   of one of the polypeptides of SEQ ID NOs: 141-241 and 378-513. In a preferred embodiment, the isolated or purified

polypeptide comprises a signal peptide of one of the polypeptides of SEQ ID NOs: 141, 143-147, 149, 150, 152, 154,

158, 161, 163-174, 177-179, 181-184, 186-189, 191, 194-196, 198, 200-203, 205, 208-226, 228, 229, 231, 233,

235-241, and 378-513.

Yet another embodiment of the present invention is an isolated or purified polypeptide comprising a mature

10   protein of one of the polypeptides of SEQ ID NOs: 141-241 and 378-513. In a preferred embodiment, the isolated or

purified polypeptide comprises a mature protein of one of the polypeptides of SEQ ID NOs: 141-145, 147, 149, 150,

152-154, 156, 157, 159-172, 176-179, 181-189, 191, 194, 195, 198, 200-226, 228, 233, 234, 236-239, 241 and

378-513.

A further embodiment of the present invention is a method of making a protein comprising one of the

15   sequences of SEQ ID NO: 141-241 and 378-513, comprising the steps of obtaining a cDNA comprising one of the

sequences of sequence of SEQ ID NO: 40-140 and 242-377, inserting the cDNA in an expression vector such that the

cDNA is operably linked to a promoter, and introducing the expression vector into a host cell whereby the host cell

produces the protein encoded by said cDNA. In one aspect of this embodiment, the method further comprises the step of

isolating the protein.

20   Another embodiment of the present invention is a protein obtainable by the method described in the preceding

paragraph.

Another embodiment of the present invention is a method of making a protein comprising the amino acid

sequence of the mature protein contained in one of the sequences of SEQ ID NO: 141-241 and 378-513, comprising the

steps of obtaining a cDNA comprising one of the nucleotides sequence of sequence of SEQ ID NO: 40-140 and 242-377

25   which encode for the mature protein, inserting the cDNA in an expression vector such that the cDNA is operably linked

to a promoter, and introducing the expression vector into a host cell whereby the host cell produces the mature protein

encoded by the cDNA. In one aspect of this embodiment, the method further comprises the step of isolating the protein.

Another embodiment of the present invention is a mature protein obtainable by the method described in the

30   preceding paragraph.

In a preferred embodiment, the above method comprises a method of making a protein comprising the amino

acid sequence of the mature protein contained in one of the sequences of SEQ ID NO: 141-145, 147, 149, 150, 152-

154, 156, 157, 159-172, 176-179, 181-189, 191, 194, 195, 198, 200-226, 228, 233, 234, 236-239, 241 and 378-

513, comprising the steps of obtaining a cDNA comprising one of the nucleotides sequence of sequence of SEQ ID NO:

-9-

40-44, 46, 48, 49, 51-53, 55, 56, 58-72, 75-78, 80-88, 90, 93, 94, 97, 99-125, 127, 132, 133, 135-138, 140, and

242-377 which encode for the mature protein, inserting the cDNA in an expression vector such that the cDNA is

operably linked to a promoter, and introducing the expression vector into a host cell whereby the host cell produces the

mature protein encoded by the cDNA. In one aspect of this embodiment, the method further comprises the step of

5   isolating the protein.

Another embodiment of the present invention is a host cell containing the purified or isolated nucleic acids

comprising the sequence of one of SEQ ID NOs: 40-140 and 242-377 or a sequence complementary thereto described

herein.

Another embodiment of the present invention is a host cell containing the purified or isolated nucleic acids

10   comprising the full coding sequences of one of SEQ ID NOs: 40-140 and 242-377, wherein the full coding sequence

comprises the sequence encoding signal peptide and the sequence encoding mature protein described herein.

Another embodiment of the present invention is a host cell containing the purified or isolated nucleic acids

comprising the nucleotides of one of SEQ ID NOs: 40-140 and 242-377 which encode a mature protein which are

described herein. Preferably, the host cell contains the purified or isolated nucleic acids comprising the nucleotides of

15   one of SEQ ID NOs: 40-44, 46, 48, 49, 51-53, 55, 56, 58-72, 75-78, 80-88, 90, 93, 94, 97, 99-125, 127, 132, 133,

135-138, 140, and 242-377 which encode a mature protein.

Another embodiment of the present invention is a host cell containing the purified or isolated nucleic acids

comprising the nucleotides of one of SEQ ID NOs: 40-140 and 242-377 which encode the signal peptide which are

described herein. Preferably, the host cell contains the purified or isolated nucleic acids comprising the nucleotides of

20   one of SEQ ID Nos.: 40, 42-46, 48, 49, 51, 53, 57, 60, 62-73, 76-78, 80-83, 85-88, 90, 93-95, 97, 99-102, 104, 107-

125, 127, 128, 130, 132, 134-140 and 242-377 which encode the signal peptide.

Another embodiment of the present invention is a purified or isolated antibody capable of specifically binding to

a protein having the sequence of one of SEQ ID NOs: 141-241 and 378-513. In one aspect of this embodiment, the

antibody is capable of binding to a polypeptide comprising at least 10 consecutive amino acids of the sequence of one of

25   SEQ ID NOs: 141-241 and 378-513.

Another embodiment of the present invention is an array of cDNAs or fragments thereof of at least 15

nucleotides in length which includes at least one of the sequences of SEQ ID NOs: 40-140 and 242-377, or one of the

sequences complementary to the sequences of SEQ ID NOs: 40-140 and 242-377, or a fragment thereof of at least 15

consecutive nucleotides. In one aspect of this embodiment, the array includes at least two of the sequences of SEQ ID

30   NOs: 40-140 and 242-377, the sequences complementary to the sequences of SEQ ID NOs: 40-140 and 242-377, or

fragments thereof of at least 15 consecutive nucleotides. In another aspect of this embodiment, the array includes at

least five of the sequences of SEQ ID NOs: 40-140 and 242-377, the sequences complementary to the sequences of

SEQ ID NOs: 40-140 and 242-377, or fragments thereof of at least 15 consecutive nucleotides.

·10·

A further embodiment of the invention encompasses purified polynucleotides comprising an insert from a clone

deposited in a deposit having an accession number selected from the group consisting of the accession numbers listed in

Table VI or a fragment thereof comprising a contiguous span of at least 8, 10, 12, 15, 20, 25, 40, 60, 100, or 200

nucleotides of said insert. An additional embodiment of the invention encompasses purified polypeptides which

5   comprise, consist of, or consist essentially of an amino acid sequence encoded by the insert from a clone deposited in a

deposit having an accession number selected from the group consisting of the accession numbers listed in Table VI, as

well as polypeptides which comprise a fragment of said amino acid sequence consisting of a signal peptide, a mature

protein, or a contiguous span of at least 5, 8, 10, 12, 15, 20, 25, 40, 60, 100, or 200 amino acids encoded by said

insert.

10   An additional embodiment of the invention encompasses purified polypeptides which comprise a contiguous

span of at least 5, 8, 10, 12, 15, 20, 25, 40, 60, 100, or 200 amino acids of SEQ ID NOs: 158, 174, 175, 196, 226,

231, 232, wherein said contiguous span comprises at least one of the amino acid positions which was not shown to be

identical to a public sequence in any of Figures 11 to 15. Also encompassed by the invention are purified

polynuculeotides encoding said polypeptides.

15

Brief Description of the Drawings

Figure 1 is a summary of a procedure for obtaining cDNAs which have been selected to include the 5' ends of

the mRNAs from which they are derived.

Figure 2 is an analysis of the 43 amino terminal amino acids of all human SwissProt proteins to determine the

20   frequency of false positives and false negatives using the techniques for signal peptide identification described herein.

Figure 3 shows the distribution of von Heijne scores for 5' ESTs in each of the categories described herein and

the probability that these 5' ESTs encode a signal peptide.

Figure 4 shows the distribution of 5' ESTs in each category and the number of 5' ESTs in each category having

a given minimum von Heijne's score.

25   Figure 5 shows the tissues from which the mRNAs corresponding to the 5' ESTs in each of the categories

described herein were obtained.

Figure 6 illustrates a method for obtaining extended cDNAs.

Figure 7 is a map of pED6dpc2. pED6dpc2 is derived from pED6dpc1 by insertion of a new polylinker to

facilitate cDNA cloning. SSt cDNAs are cloned between EcoRI and NotI. PED vectors are described in Kaufman et al.

30   (1991), NAR 19: 4485-4490.

Figure 8 provides a schematic description of the promoters isolated and the way they are assembled with the

corresponding 5' tags.

Figure 9 describes the transcription factor binding sites present in each of these promoters.

·11·

Figure 10 is an alignment of the protein of SEQ ID NO: 217 with the human protein TFAR19 that may play a role in apoptosis (Genbank accession number AF014955, SEQ ID NO: 516).

Figure 11 is an alignment of the proteins of SEQ ID NOs: 174, 175 and 232 with a human secreted protein (Genseq accession number W36955, SEQ ID NO: 517).

5   Figure 12 is an alignment of the protein of SEQ ID NO: 231 with the human E25 protein (Genbank accession number AF038953, SEQ ID NO: 515).

Figure 13 is an alignment of the protein of SEQ ID NO: 196 with the human seventransmembrane protein (Genbank accession number Y11395, SEQ ID NO: 518).

Figure 14 is an alignment of the protein of SEQ ID NOs: 158 with the murine subunit 7a of the COP9 complex

10 (Genbank accession number AF071316, SEQ ID NO: 519).

Figure 15 is an alignment of the protein of SEQ ID NO: 226 with the bovine subunit B14.5B of the NADH-ubiquinone oxidureductase complex (Arizmendi *et al, FEBS Lett.*, 313 : 80-84 (1992) and Swissprot accession ·number Q02827, SEQ ID NO: 514).

Detailed Description of the Preferred Embodiment

15 I. Obtaining 5′ ESTs

The present extended cDNAs were obtained using 5′ ESTs which were isolated as described below.

A. Chemical Methods for Obtaining mRNAs having Intact 5′ Ends

In order to obtain the 5′ ESTs used to obtain the extended cDNAs of the present invention, mRNAs having intact 5′ ends must be obtained. Currently, there are two approaches for obtaining such mRNAs. One of these

20 approaches is a chemical modification method involving derivatization of the 5′ ends of the mRNAs and selection of the derivatized mRNAs. The 5′ ends of eucaryotic mRNAs possess a structure referred to as a "cap" which comprises a guanosine methylated at the 7 position. The cap is joined to the first transcribed base of the mRNA by a 5′, 5′-triphosphate bond. In some instances, the 5′ guanosine is methylated in both the 2 and 7 positions. Rarely, the 5′ guanosine is trimethylated at the 2, 7 and 7 positions. In the chemical method for obtaining mRNAs having intact 5′

25 ends, the 5′ cap is specifically derivatized and coupled to a reactive group on an immobilizing substrate. This specific derivatization is based on the fact that only the ribose linked to the methylated guanosine at the 5′ end of the mRNA and the ribose linked to the base at the 3′ terminus of the mRNA, possess 2′, 3′-cis diols. Optionally, where the 3′ terminal ribose has a 2′, 3′-cis diol, the 2′, 3′-cis diol at the 3′ end may be chemically modified, substituted, converted, or eliminated, leaving only the ribose linked to the methylated guanosine at the 5′ end of the mRNA with a 2′, 3′-cis diol. A

30 variety of techniques are available for eliminating the 2′, 3′-cis diol on the 3′ terminal ribose. For example, controlled alkaline hydrolysis may be used to generate mRNA fragments in which the 3′ terminal ribose is a 3′-phosphate, 2′-phosphate or (2′, 3′)-cyclophosphate. Thereafter, the fragment which includes the original 3′ ribose may be eliminated from the mixture through chromatography on an oligo-dT column. Alternatively, a base which lacks the 2′, 3′-cis diol

-12-

may be added to the 3' end of the mRNA using an RNA ligase such as T4 RNA ligase. Example 1 below describes a method for ligation of pCp to the 3' end of messenger RNA.

EXAMPLE 1

Ligation of the Nucleoside Diphosphate pCp to the 3' End of Messenger RNA

5   1 μg of RNA was incubated in a final reaction medium of 10 μl in the presence of 5 U of T_4 phage RNA ligase in the buffer provided by the manufacturer (Gibco - BRL), 40 U of the RNase inhibitor RNasin (Promega) and, 2 μl of ^{32}pCp (Amersham #PB 10208).

The incubation was performed at 37°C for 2 hours or overnight at 7-8°C.

Following modification or elimination of the 2', 3'-cis diol at the 3' ribose, the 2', 3'-cis diol present at the 5'
10 end of the mRNA may be oxidized using reagents such as $NaBH_4$, $NaBH_3CN$, or sodium periodate, thereby converting the 2', 3'-cis diol to a dialdehyde. Example 2 describes the oxidation of the 2', 3'-cis diol at the 5' end of the mRNA with sodium periodate.

EXAMPLE 2

Oxidation of 2', 3'-cis diol at the 5' End of the mRNA

15   0.1 OD unit of either a capped oligoribonucleotide of 47 nucleotides (including the cap) or an uncapped oligoribonucleotide of 46 nucleotides were treated as follows. The oligoribonucleotides were produced by in vitro transcription using the transcription kit "AmpliScribe T7" (Epicentre Technologies). As indicated below, the DNA template for the RNA transcript contained a single cytosine. To synthesize the uncapped RNA, all four NTPs were included in the in vitro transcription reaction. To obtain the capped RNA, GTP was replaced by an analogue of the cap,
20 m7G(5')ppp(5')G. This compound, recognized by polymerase, was incorporated into the 5' end of the nascent transcript during the step of initiation of transcription but was not capable of incorporation during the extension step. Consequently, the resulting RNA contained a cap at its 5' end. The sequences of the oligoribonucleotides produced by the in vitro transcription reaction were:

+Cap:

25 5'm7GpppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3' (SEQ ID NO:1)

-Cap:

5'-pppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3' (SEQ ID NO:2)

The oligoribonucleotides were dissolved in 9 μl of acetate buffer (0.1 M sodium acetate, pH 5.2) and 3 μl of freshly prepared 0.1 M sodium periodate solution. The mixture was incubated for 1 hour in the dark at 4°C or room
30 temperature. Thereafter, the reaction was stopped by adding 4 μl of 10% ethylene glycol. The product was ethanol precipitated, resuspended in 10μl or more of water or appropriate buffer and dialyzed against water.

The resulting aldehyde groups may then be coupled to molecules having a reactive amine group, such as hydrazine, carbazide, thiocarbazide or semicarbazide groups, in order to facilitate enrichment of the 5' ends of the mRNAs. Molecules having reactive amine groups which are suitable for use in selecting mRNAs having intact 5' ends

include avidin, proteins, antibodies, vitamins, ligands capable of specifically binding to receptor molecules, or

oligonucleotides. Example 3 below describes the coupling of the resulting dialdehyde to biotin.

EXAMPLE 3

Coupling of the Dialdehyde with Biotin

The oxidation product obtained in Example 2 was dissolved in 50 μl of sodium acetate at a pH of between 5

and 5.2 and 50 μl of freshly prepared 0.02 M solution of biotin hydrazide in a methoxyethanol/water mixture (1:1) of

formula:
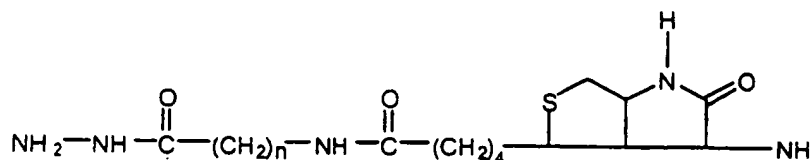
In the compound used in these experiments, n = 5. However, it will be appreciated that other commercially

available hydrazides may also be used, such as molecules of the formula above in which n varies from 0 to 5.

The mixture was then incubated for 2 hours at 37°C. Following the incubation, the mixture was precipitated

with ethanol and dialyzed against distilled water.

Example 4 demonstrates the specificity of the biotinylation reaction.

EXAMPLE 4

Specificity of Biotinylation

The specificity of the biotinylation for capped mRNAs was evaluated by gel electrophoresis of the following

samples:

Sample 1. The 46 nucleotide uncapped in vitro transcript prepared as in Example 2 and labeled with ^{32}pCp as

described in Example 1.

Sample 2. The 46 nucleotide uncapped in vitro transcript prepared as in Example 2, labeled with ^{32}pCp as

described in Example 1, treated with the oxidation reaction of Example 2, and subjected to the biotinylation conditions of

Example 3.

Sample 3. The 47 nucleotide capped in vitro transcript prepared as in Example 2 and labeled with ^{32}pCp as

described in Example 1.

Sample 4. The 47 nucleotide capped in vitro transcript prepared as in Example 2, labeled with ^{32}pCp as

described in Example 1, treated with the oxidation reaction of Example 2, and subjected to the biotinylation conditions of

Example 3.

Samples 1 and 2 had indentical migration rates, demonstrating that the uncapped RNAs were not oxidized and

biotinylated. Sample 3 migrated more slowly than Samples 1 and 2, while Sample 4 exhibited the slowest migration.

·14·

The difference in migration of the RNAs in Samples 3 and 4 demonstrates that the capped RNAs were specifically biotinylated.

In some cases, mRNAs having intact 5' ends may be enriched by binding the molecule containing a reactive amine group to a suitable solid phase substrate such as the inside of the vessel containing the mRNAs, magnetic beads,

5 chromatography matrices, or nylon or nitrocellulose membranes. For example, where the molecule having a reactive amine group is biotin, the solid phase substrate may be coupled to avidin or streptavidin. Alternatively, where the molecule having the reactive amine group is an antibody or receptor ligand, the solid phase substrate may be coupled to the cognate antigen or receptor. Finally, where the molecule having a reactive amine group comprises an oligonucleotide, the solid phase substrate may comprise a complementary oligonucleotide.

10   The mRNAs having intact 5' ends may be released from the solid phase following the enrichment procedure. For example, where the dialdehyde is coupled to biotin hydrazide and the solid phase comprises streptavidin, the mRNAs may be released from the solid phase by simply heating to 95 degrees Celsius in 2% SDS. In some methods, the molecule having a reactive amine group may also be cleaved from the mRNAs having intact 5' ends following enrichment. Example 5 describes the capture of biotinylated mRNAs with streptavidin coated beads and the release of the

15 biotinylated mRNAs from the beads following enrichment.

EXAMPLE 5

Capture and Release of Biotinylated mRNAs Using Strepatividin Coated Beads

The streptavidin-coated magnetic beads were prepared according to the manufacturer's instructions (CPG Inc., USA). The biotinylated mRNAs were added to a hybridization buffer (1.5 M NaCl, pH 5 - 6). After incubating for 30

20 minutes, the unbound and nonbiotinylated material was removed. The beads were washed several times in water with 1% SDS. The beads obtained were incubated for 15 minutes at 95°C in water containing 2% SDS.

Example 6 demonstrates the efficiency with which biotinylated mRNAs were recovered from the streptavidin coated beads.

EXAMPLE 6

25   #### Efficiency of Recovery of Biotinylated mRNAs

The efficiency of the recovery procedure was evaluated as follows. RNAs were labeled with ^{32}pCp, oxidized, biotinylated and bound to streptavidin coated beads as described above. Subsequently, the bound RNAs were incubated for 5, 15 or 30 minutes at 95°C in the presence of 2% SDS.

The products of the reaction were analyzed by electrophoresis on 12% polyacrylamide gels under denaturing

30 conditions (7 M urea). The gels were subjected to autoradiography. During this manipulation, the hydrazone bonds were not reduced.

Increasing amounts of nucleic acids were recovered as incubation times in 2% SDS increased, demonstrating that biotinylated mRNAs were efficiently recovered.

·15·

In an alternative method for obtaining mRNAs having intact 5' ends, an oligonucleotide which has been derivatized to contain a reactive amine group is specifically coupled to mRNAs having an intact cap. Preferably, the 3' end of the mRNA is blocked prior to the step in which the aldehyde groups are joined to the derivatized oligonucleotide, as described above, so as to prevent the derivatized oligonucleotide from being joined to the 3' end of the mRNA. For

5   example, pCp may be attached to the 3' end of the mRNA using T4 RNA ligase. However, as discussed above, blocking the 3' end of the mRNA is an optional step. Derivatized oligonucleotides may be prepared as described below in Example 7.

EXAMPLE 7

Derivatization of the Oligonucleotide

10   An oligonucleotide phosphorylated at its 3' end was converted to a 3' hydrazide in 3' by treatment with an aqueous solution of hydrazine or of dihydrazide of the formula $H_2N(R1)NH_2$ at about 1 to 3 M, and at pH 4.5, in the presence of a carbodiimide type agent soluble in water such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide at a final concentration of 0.3 M at a temperature of 8°C overnight.

The derivatized oligonucleotide was then separated from the other agents and products using a standard

15   technique for isolating oligonucleotides.

As discussed above, the mRNAs to be enriched may be treated to eliminate the 3' OH groups which may be present thereon. This may be accomplished by enzymatic ligation of sequences lacking a 3' OH, such as pCp, as described above in Example 1. Alternatively, the 3' OH groups may be eliminated by alkaline hydrolysis as described in Example 8 below.

20   · EXAMPLE 8

Alkaline Hydrolysis of mRNA

The mRNAs may be treated with alkaline hydrolysis as follows. In a total volume of 100µl of 0.1N sodium hydroxide, 1.5µg mRNA is incubated for 40 to 60 minutes at 4°C. The solution is neutralized with acetic acid and precipitated with ethanol.

25   Following the optional elimination of the 3' OH groups, the diol groups at the 5' ends of the mRNAs are oxidized as described below in Example 9.

EXAMPLE 9

Oxidation of Diols

Up to 1 OD unit of RNA was dissolved in 9 µl of buffer (0.1 M sodium acetate, pH 6-7 or water) and 3 µl of

30   freshly prepared 0.1 M sodium periodate solution. The reaction was incubated for 1 h in the dark at 4°C or room temperature. Following the incubation, the reaction was stopped by adding 4 µl of 10% ethylene glycol. Thereafter the mixture was incubated at room temperature for 15 minutes. After ethanol precipitation, the product was resuspended in 10µl or more of water or appropriate buffer and dialyzed against water.

-16-

Following oxidation of the diol groups at the 5' ends of the mRNAs, the derivatized oligonucleotide was joined

to the resulting aldehydes as described in Example 10.

EXAMPLE 10

Reaction of Aldehydes with Derivatized Oligonucleotides

5   The oxidized mRNA was dissolved in an acidic medium such as 50 µl of sodium acetate pH 4-6. 50 µl of a

solution of the derivatized oligonucleotide was added such that an mRNA:derivatized oligonucleotide ratio of 1:20 was

obtained and mixture was reduced with a borohydride. The mixture was allowed to incubate for 2 h at 37°C or

overnight (14 h) at 10°C. The mixture was ethanol precipitated, resuspended in 10µl or more of water or appropriate

buffer and dialyzed against distilled water. If desired, the resulting product may be analyzed using acrylamide gel

10   electrophoresis, HPLC analysis, or other conventional techniques.

Following the attachment of the derivatized oligonucleotide to the mRNAs, a reverse transcription reaction may

be performed as described in Example 11 below.

EXAMPLE 11

Reverse Transcription of mRNAs

15   An oligodeoxyribonucleotide was derivatized as follows. 3 OD units of an oligodeoxyribonucleotide of sequence

ATCAAGAATTCGCACGAGACCATTA (SEQ ID NO:3) having 5'-OH and 3'-P ends were dissolved in 70 µl of a 1.5 M

hydroxybenzotriazole solution, pH 5.3, prepared in dimethylformamide/water (75:25) containing 2 µg of 1-ethyl-3-(3-

dimethylaminopropyl)carbodiimide. The mixture was incubated for 2 h 30 min at 22°C. The mixture was then

precipitated twice in LiClO₄/acetone. The pellet was resuspended in 200 µl of 0.25 M hydrazine and incubated at 8°C

20   from 3 to 14 h. Following the hydrazine reaction, the mixture was precipitated twice in LiClO₄/acetone.

The messenger RNAs to be reverse transcribed were extracted from blocks of placenta having sides of 2 cm

which had been stored at -80°C. The mRNA was extracted using conventional acidic phenol techniques. Oligo-dT

chromatography was used to purify the mRNAs. The integrity of the mRNAs was checked by Northern-blotting.

The diol groups on 7 µg of the placental mRNAs were oxidized as described above in Example 9. The

25   derivatized oligonucleotide was joined to the mRNAs as described in Example 10 above except that the precipitation step

was replaced by an exclusion chromatography step to remove derivatized oligodeoxyribonucleotides which were not

joined to mRNAs. Exclusion chromatography was performed as follows:

10 ml of AcA34 (BioSepra#230151) gel were equilibrated in 50 ml of a solution of 10 mM Tris pH 8.0, 300

mM NaCl, 1 mM EDTA, and 0.05% SDS. The mixture was allowed to sediment. The supernatant was eliminated and

30   the gel was resuspended in 50 ml of buffer. This procedure was repeated 2 or 3 times.

A glass bead (diameter 3 mm) was introduced into a 2 ml disposable pipette (length 25 cm). The pipette was

filled with the gel suspension until the height of the gel stabilized at 1 cm from the top of the pipette. The column was

then equilibrated with 20 ml of equilibration buffer (10 mM Tris HCl pH 7.4, 20 mM NaCl).

10 μl of the mRNA which had been reacted with the derivatized oligonucleotide were mixed in 39 μl of 10 mM urea and 2 μl of blue-glycerol buffer, which had been prepared by dissolving 5 mg of bromophenol blue in 60% glycerol (v/v), and passing the mixture through a filter with a filter of diameter 0.45 μm.

The column was loaded. As soon as the sample had penetrated, equilibration buffer was added. 100 μl
5   fractions were collected. Derivatized oligonucleotide which had not been attached to mRNA appeared in fraction 16 and later fractions. Fractions 3 to 15 were combined and precipitated with ethanol.

The mRNAs which had been reacted with the derivatized oligonucleotide were spotted on a nylon membrane and hybridized to a radioactive probe using conventional techniques. The radioactive probe used in these hybridizations was an oligodeoxyribonucleotide of sequence TAATGGTCTCGTGCGAATTCTTGAT (SEQ ID NO:4) which was
10 anticomplementary to the derivatized oligonucleotide and was labeled at its 5' end with ^{32}P. 1/10th of the mRNAs which had been reacted with the derivatized oligonucleotide was spotted in two spots on the membrane and the membrane was visualized by autoradiography after hybridization of the probe. A signal was observed, indicating that the derivatized oligonucleotide had been joined to the mRNA.

The remaining 9/10 of the mRNAs which had been reacted with the derivatized oligonucleotide was reverse
15 transcribed as follows. A reverse transcription reaction was carried out with reverse transcriptase following the manufacturer's instructions. To prime the reaction, 50 pmol of nonamers with random sequence were used.

A portion of the resulting cDNA was spotted on a positively charged nylon membrane using conventional methods. The cDNAs were spotted on the membrane after the cDNA:RNA heteroduplexes had been subjected to an alka-line hydrolysis in order to eliminate the RNAs. An oligonucleotide having a sequence identical to that of the derivatized
20 oligonucleotide was labeled at its 5' end with ^{32}P and hybridized to the cDNA blots using conventional techniques. Single-stranded cDNAs resulting from the reverse transcription reaction were spotted on the membrane. As controls, the blot contained 1 pmol, 100 fmol, 50 fmol, 10 fmol and 1 fmol respectively of a control oligodeoxyribonucleotide of sequence identical to that of the derivatized oligonucleotide. The signal observed in the spots containing the cDNA indicated that approximately 15 fmol of the derivatized oligonucleotide had been reverse transcribed.

25 These results demonstrate that the reverse transcription can be performed through the cap and, in particular, that reverse transcriptase crosses the 5'-P-P-P-5' bond of the cap of eukaryotic messenger RNAs.

The single stranded cDNAs obtained after the above first strand synthesis were used as template for PCR reactions. Two types of reactions were carried out. First, specific amplification of the mRNAs for the alpha globin, dehydrogenase, pp15 and elongation factor E4 were carried out using the following pairs of oligodeoxyribonucleotide
30 primers.

alpha-globin

GLO-S: CCG ACA AGA CCA ACG TCA AGG CCG C (SEQ ID NO:5)

GLO-As: TCA CCA GCA GGC AGT GGC TTA GGA G 3' (SEQ ID NO:6)

dehydrogenase

-18-

3 DH-S: AGT GAT TCC TGC TAC TTT GGA TGG C (SEQ ID NO:7)

3 DH-As: GCT TGG TCT TGT TCT GGA GTT TAG A (SEQ ID NO:8)

                        pp15

PP15-S: TCC AGA ATG GGA GAC AAG CCA ATT T (SEQ ID NO:9)

5   PP15-As: AGG GAG GAG GAA ACA GCG TGA GTC C (SEQ ID NO:10)

                Elongation factor E4

EFA1-S: ATG GGA AAG GAA AAG ACT CAT ATC A (SEQ ID NO:11)

EF1A-As: AGC AGC AAC AAT CAG GAC AGC ACA G (SEQ ID NO:12)

        Non specific amplifications were also carried out with the antisense (_As) oligodeoxyribonucleotides of the

10   pairs described above and a primer chosen from the sequence of the derivatized oligodeoxyribonucleotide

(ATCAAGAATTCGCACGAGACCATTA) (SEQ ID NO:13).

        A 1.5% agarose gel containing the following samples corresponding to the PCR products of reverse

transcription was stained with ethidium bromide. (1/20th of the products of reverse transcription were used for each

PCR reaction).

15   Sample 1: The products of a PCR reaction using the globin primers of SEQ ID NOs 5 and 6 in the presence of

cDNA.

        Sample 2: The products of a PCR reaction using the globin primers of SEQ ID NOs 5 and 6 in the absence of

added cDNA.

        Sample 3: The products of a PCR reaction using the dehydrogenase primers of SEQ ID NOs 7 and 8 in the

20   presence of cDNA.

        Sample 4: The products of a PCR reaction using the dehydrogenase primers of SEQ ID NOs 7 and 8 in the

absence of added cDNA.

        Sample 5: The products of a PCR reaction using the pp15 primers of SEQ ID NOs 9 and 10 in the presence of

cDNA.

25   Sample 6: The products of a PCR reaction using the pp15 primers of SEQ ID NOs 9 and 10 in the absence of

added cDNA.

        Sample 7: The products of a PCR reaction using the ElE4 primers of SEQ ID NOs 11 and 12 in the presence of

added cDNA.

        Sample 8: The products of a PCR reaction using the ElE4 primers of SEQ ID NOs 11 and 12 in the absence of

30   added cDNA.

        In Samples 1, 3, 5 and 7, a band of the size expected for the PCR product was observed, indicating the

presence of the corresponding sequence in the cDNA population.

        PCR reactions were also carried out with the antisense oligonucleotides of the globin and dehydrogenase

primers (SEQ ID NOs 6 and 8) and an oligonucleotide whose sequence corresponds to that of the derivatized

oligonucleotide. The presence of PCR products of the expected size in the samples corresponding to samples 1 and 3 above indicated that the derivatized oligonucleotide had been incorporated.

The above examples summarize the chemical procedure for enriching mRNAs for those having intact 5' ends. Further detail regarding the chemical approaches for obtaining mRNAs having intact 5' ends are disclosed in

5   International Application No. WO96/34981, published November 7, 1996.

Strategies based on the above chemical modifications to the 5' cap structure may be utilized to generate cDNAs which have been selected to include the 5' ends of the mRNAs from which they are derived. In one version of such procedures, the 5' ends of the mRNAs are modified as described above. Thereafter, a reverse transcription reaction is conducted to extend a primer complementary to the mRNA to the 5' end of the mRNA. Single stranded RNAs

10   are eliminated to obtain a population of cDNA/mRNA heteroduplexes in which the mRNA includes an intact 5' end. The resulting heteroduplexes may be captured on a solid phase coated with a molecule capable of interacting with the molecule used to derivatize the 5' end of the mRNA. Thereafter, the strands of the heteroduplexes are separated to recover single stranded first cDNA strands which include the 5' end of the mRNA. Second strand cDNA synthesis may then proceed using conventional techniques. For example, the procedures disclosed in WO 96/34981 or in Carninci, P. et

15   al. High-Efficiency Full-Length cDNA Cloning by Biotinylated CAP Trapper. Genomics 37:327-336 (1996) may be employed to select cDNAs which include the sequence derived from the 5' end of the coding sequence of the mRNA.

Following ligation of the oligonucleotide tag to the 5' cap of the mRNA, a reverse transcription reaction is conducted to extend a primer complementary to the mRNA to the 5' end of the mRNA. Following elimination of the RNA component of the resulting heteroduplex using standard techniques, second strand cDNA synthesis is conducted with a

20   primer complementary to the oligonucleotide tag.

Figure 1 summarizes the above procedures for obtaining cDNAs which have been selected to include the 5' ends of the mRNAs from which they are derived.

B. Enzymatic Methods for Obtaining mRNAs having Intact 5' Ends

Other techniques for selecting cDNAs extending to the 5' end of the mRNA from which they are derived are

25   fully enzymatic. Some versions of these techniques are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc complets: difficultes et perspectives nouvelles. Apports pour l'etude de la regulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EPO 625572 and Kato et al. Construction of a Human Full-Length cDNA Bank. Gene 150:243-250 (1994).

Briefly, in such approaches, isolated mRNA is treated with alkaline phosphatase to remove the phosphate

30   groups present on the 5' ends of uncapped incomplete mRNAs. Following this procedure, the cap present on full length mRNAs is enzymatically removed with a decapping enzyme such as T4 polynucleotide kinase or tobacco acid pyrophosphatase. An oligonucleotide, which may be either a DNA oligonucleotide or a DNA-RNA hybrid oligonucleotide having RNA at its 3' end, is then ligated to the phosphate present at the 5' end of the decapped mRNA using T4 RNA

-20-

ligase. The oligonucleotide may include a restriction site to facilitate cloning of the cDNAs following their synthesis. Example 12 below describes one enzymatic method based on the doctoral thesis of Dumas.

EXAMPLE 12

Enzymatic Approach for Obtaining 5' ESTs

5   Twenty micrograms of PolyA+ RNA were dephosphorylated using Calf Intestinal Phosphatase (Biolabs). After a phenol chloroform extraction, the cap structure of mRNA was hydrolysed using the Tobacco Acid Pyrophosphatase (purified as described by Shinshi et al., Biochemistry 15: 2185-2190, 1976) and a hemi 5'DNA/RNA-3' oligonucleotide having an unphosphorylated 5' end, a stretch of adenosine ribophosphate at the 3' end, and an EcoRI site near the 5' end was ligated to the 5'P ends of mRNA using the T4 RNA ligase (Biolabs). Oligonucleotides suitable for use in this

10 procedure are preferably 30-50 bases in length. Oligonucleotides having an unphosphorylated 5' end may be synthesized by adding a fluorochrome at the 5' end. The inclusion of a stretch of adenosine ribophosphates at the 3' end of the oligonucleotide increases ligation efficiency. It will be appreciated that the oligonucleotide may contain cloning sites other than EcoRI.

       Following ligation of the oligonucleotide to the phosphate present at the 5' end of the decapped mRNA, first

15 and second strand cDNA synthesis may be carried out using conventional methods or those specified in EPO 625,572 and Kato et al. Construction of a Human Full-Length cDNA Bank. Gene 150:243-250 (1994), and Dumas Milne Edwards, *supra*. The resulting cDNA may then be ligated into vectors such as those disclosed in Kato et al. Construction of a Human Full-Length cDNA Bank. Gene 150:243-250 (1994) or other nucleic acid vectors known to those skilled in the art using techniques such as those described in Sambrook et al., Molecular Cloning: A Laboratory Manual 2d Ed., Cold

20 Spring Harbor Laboratory Press, 1989.

II. Characterization of 5' ESTs

       The above chemical and enzymatic approaches for enriching mRNAs having intact 5' ends were employed to obtain 5' ESTs. First, mRNAs were prepared as described in Example 13 below.

EXAMPLE 13

25   ### Preparation of mRNA

       Total human RNAs or PolyA+ RNAs derived from 29 different tissues were respectively purchased from LABIMO and CLONTECH and used to generate 44 cDNA libraries as described below. The purchased RNA had been isolated from cells or tissues using acid guanidium thiocyanate-phenol-chloroform extraction (Chomczyniski, P and Sacchi, N., Analytical Biochemistry 162:156-159, 1987). PolyA+ RNA was isolated from total RNA (LABIMO) by

30 two passes of oligodT chromatography, as described by Aviv and Leder (Aviv, H. and Leder, P., Proc. Natl. Acad. Sci. USA 69:1408-1412, 1972) in order to eliminate ribosomal RNA.

       The quality and the integrity of the poly A+ were checked. Northern blots hybridized with a globin probe were used to confirm that the mRNAs were not degraded. Contamination of the PolyA+ mRNAs by ribosomal sequences was checked using RNAs blots and a probe derived from the sequence of the 28S RNA. Preparations of mRNAs with less

than 5% of ribosomal RNAs were used in library construction. To avoid constructing libraries with RNAs contaminated

by exogenous sequences (prokaryotic or fungal), the presence of bacterial 16S ribosomal sequences or of two highly

expressed mRNAs was examined using PCR.

Following preparation of the mRNAs, the above described chemical and/or the enzymatic procedures for

5 enriching mRNAs having intact 5' ends discussed above were employed to obtain 5' ESTs from various tissues. In both

approaches an oligonucleotide tag was attached to the cap at the 5' ends of the mRNAs. The oligonucleotide tag had an

EcoRI site therein to facilitate later cloning procedures.

Following attachment of the oligonucleotide tag to the mRNA by either the chemical or enzymatic methods, the

integrity of the mRNA was examined by performing a Northern blot with 200-500ng of mRNA using a probe

10 complementary to the oligonucleotide tag.

EXAMPLE 14

cDNA Synthesis Using mRNA Templates Having Intact 5' Ends

For the mRNAs joined to oligonucleotide tags using both the chemical and enzymatic methods, first strand

cDNA synthesis was performed using reverse transcriptase with random nonamers as primers. In order to protect

15 internal EcoRI sites in the cDNA from digestion at later steps in the procedure, methylated dCTP was used for first

strand synthesis. After removal of RNA by an alkaline hydrolysis, the first strand of cDNA was precipitated using

isopropanol in order to eliminate residual primers.

For both the chemical and the enzymatic methods, the second strand of the cDNA was synthesized with a

Klenow fragment using a primer corresponding to the 5'end of the ligated oligonucleotide described in Example 12.

20 Preferably, the primer is 20-25 bases in length. Methylated dCTP was also used for second strand synthesis in order to

protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following cDNA synthesis, the cDNAs were cloned into pBlueScript as described in Example 15 below.

EXAMPLE 15

Insertion of cDNAs into BlueScript

25 Following second strand synthesis, the ends of the cDNA were blunted with T4 DNA polymerase (Biolabs) and

the cDNA was digested with EcoRI. Since methylated dCTP was used during cDNA synthesis, the EcoRI site present in

the tag was the only site which was hemi-methylated. Consequently, only the EcoRI site in the oligonucleotide tag was

susceptible to EcoRI digestion. The cDNA was then size fractionated using exclusion chromatography (AcA, Biosepra).

Fractions corresponding to cDNAs of more than 150 bp were pooled and ethanol precipitated. The cDNA was

30 directionally cloned into the Smal and EcoRI ends of the phagemid pBlueScript vector (Stratagene). The ligation mixture

was electroporated into bacteria and propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached were selected as described in Example 16 below.

EXAMPLE 16

Selection of Clones Having the Oligonucleotide Tag Attached Thereto

-22-

The plasmid DNAs containing 5' EST libraries made as described above were purified (Qiagen). A positive selection of the tagged clones was performed as follows. Briefly, in this selection procedure, the plasmid DNA was converted to single stranded DNA using gene II endonuclease of the phage F1 in combination with an exonuclease (Chang et al., **Gene** 127:95-8, 1993) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA was

5   then purified using paramagnetic beads as described by Fry et al., **Biotechniques**, 13: 124-131, 1992. In this procedure, the single stranded DNA was hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide described in Example 13. Preferably, the primer has a length of 20-25 bases. Clones including a sequence complementary to the biotinylated oligonucleotide were captured by incubation with streptavidin coated magnetic beads followed by magnetic selection. After capture of the positive clones, the plasmid DNA was

10   released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as the ThermoSequenase obtained from Amersham Pharmacia Biotech. Alternatively, protocols such as the Gene Trapper kit (Gibco BRL) may be used. The double stranded DNA was then electroporated into bacteria. The percentage of positive clones having the 5' tag oligonucleotide was estimated to typically rank between 90 and 98% using dot blot analysis.

Following electroporation, the libraries were ordered in 384-microtiter plates (MTP). A copy of the MTP was

15   stored for future needs. Then the libraries were transferred into 96 MTP and sequenced as described below.

EXAMPLE 17

Sequencing of Inserts in Selected Clones

Plasmid inserts were first amplified by PCR on PE 9600 thermocyclers (Perkin-Elmer), using standard SETA-A and SETA-B primers (Genset SA), AmpliTaqGold (Perkin-Elmer), dNTPs (Boehringer), buffer and cycling conditions as

20   recommended by the Perkin-Elmer Corporation.

PCR products were then sequenced using automatic ABI Prism 377 sequencers (Perkin Elmer, Applied Biosystems Division, Foster City, CA). Sequencing reactions were performed using PE 9600 thermocyclers (Perkin Elmer) with standard dye-primer chemistry and ThermoSequenase (Amersham Life Science). The primers used were either T7 or 21M13 (available from Genset SA) as appropriate. The primers were labeled with the JOE, FAM, ROX and TAMRA

25   dyes. The dNTPs and ddNTPs used in the sequencing reactions were purchased from Boehringer. Sequencing buffer, reagent concentrations and cycling conditions were as recommended by Amersham.

Following the sequencing reaction, the samples were precipitated with EtOH, resuspended in formamide loading buffer, and loaded on a standard 4% acrylamide gel. Electrophoresis was performed for 2.5 hours at 3000V on an ABI 377 sequencer, and the sequence data were collected and analyzed using the ABI Prism DNA Sequencing

30   Analysis Software, version 2.1.2.

The sequence data from the 44 cDNA libraries made as described above were transferred to a proprietary database, where quality control and validation steps were performed. A proprietary base-caller ("Trace"), working using a Unix system automatically flagged suspect peaks, taking into account the shape of the peaks, the inter-peak resolution, and the noise level. The proprietary base-caller also performed an automatic trimming. Any stretch of 25 or

-23-

fewer bases having more than 4 suspect peaks was considered unreliable and was discarded. Sequences corresponding to cloning vector or ligation oligonucleotides were automatically removed from the EST sequences. However, the resulting EST sequences may contain 1 to 5 bases belonging to the above mentioned sequences at their 5' end. If needed, these can easily be removed on a case by case basis.

5   Thereafter, the sequences were transferred to the proprietary NETGENE™ Database for further analysis as described below.

          Following sequencing as described above, the sequences of the 5' ESTs were entered in a proprietary database called NETGENE™ for storage and manipulation. It will be appreciated by those skilled in the art that the data could be stored and manipulated on any medium which can be read and accessed by a computer. Computer readable media

10   include magnetically readable media, optically readable media, or electronically readable media. For example, the computer readable media may be a hard disc, a floppy disc, a magnetic tape, CD-ROM, RAM, or ROM as well as other types of other media known to those skilled in the art.

          In addition, the sequence data may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, the sequence data may be stored as text in a word processing file, such as

15   MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE.

          The computer readable media on which the sequence information is stored may be in a personal computer, a network, a server or other computer systems known to those skilled in the art. The computer or other system preferably includes the storage media described above, and a processor for accessing and manipulating the sequence data.

20   Once the sequence data has been stored it may be manipulated and searched to locate those stored sequences which contain a desired nucleic acid sequence or which encode a protein having a particular functional domain. For example, the stored sequence information may be compared to other known sequences to identify homologies, motifs implicated in biological function, or structural motifs.

          Programs which may be used to search or compare the stored sequences include the MacPattern (EMBL),

25   BLAST, and BLAST2 program series (NCBI), basic local alignment search tool programs for nucleotide (BLASTN) and peptide (BLASTX) comparisons (Altschul et al, J. Mol. Biol. 215: 403 (1990)) and FASTA (Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85: 2444 (1988)). The BLAST programs then extend the alignments on the basis of defined match and mismatch criteria.

          Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-

30   helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

·24·

Before searching the cDNAs in the NETGENE™ database for sequence motifs of interest, cDNAs derived from mRNAs which were not of interest were identified and eliminated from further consideration as described in Example 18 below.

EXAMPLE 18

5   Elimination of Undesired Sequences from Further Consideration

5' ESTs in the NETGENE™ database which were derived from undesired sequences such as transfer RNAs, ribosomal RNAs, mitochondrial RNAs, procaryotic RNAs, fungal RNAs, Alu sequences, L1 sequences, or repeat sequences were identified using the FASTA and BLASTN programs with the parameters listed in Table II.

To eliminate 5' ESTs encoding tRNAs from further consideration, the 5' EST sequences were compared to the

10   sequences of 1190 known tRNAs obtained from EMBL release 38, of which 100 were human. The comparison was performed using FASTA on both strands of the 5' ESTs. Sequences having more than 80% homology over more than 60 nucleotides were identified as tRNA. Of the 144,341 sequences screened, 26 were identified as tRNAs and eliminated from further consideration.

To eliminate 5' ESTs encoding rRNAs from further consideration, the 5' EST sequences were compared to the

15   sequences of 2497 known rRNAs obtained from EMBL release 38, of which 73 were human. The comparison was performed using BLASTN on both strands of the 5' ESTs with the parameter S = 108. Sequences having more than 80% homology over stretches longer than 40 nucleotides were identified as rRNAs. Of the 144,341 sequences screened, 3,312 were identified as rRNAs and eliminated from further consideration.

To eliminate 5' ESTs encoding mtRNAs from further consideration, the 5' EST sequences were compared to

20   the sequences of the two known mitochondrial genomes for which the entire genomic sequences are available and all sequences transcribed from these mitochondrial genomes including tRNAs, rRNAs, and mRNAs for a total of 38 sequences. The comparison was performed using BLASTN on both strands of the 5' ESTs with the parameter S = 108. Sequences having more than 80% homology over stretches longer than 40 nucleotides were identified as mtRNAs. Of the 144,341 sequences screened, 6,110 were identified as mtRNAs and eliminated from further consideration.

25   Sequences which might have resulted from exogenous contaminants were eliminated from further consideration by comparing the 5' EST sequences to release 46 of the EMBL bacterial and fungal divisions using BLASTN with the parameter S = 144. All sequences having more than 90% homology over at least 40 nucleotides were identified as exogenous contaminants. Of the 42 cDNA libraries examined, the average percentages of procaryotic and fungal sequences contained therein were 0.2% and 0.5% respectively. Among these sequences, only one could be

30   identified as a sequence specific to fungi. The others were either fungal or procaryotic sequences having homologies with vertebrate sequences or including repeat sequences which had not been masked during the electronic comparison.

In addition, the 5' ESTs were compared to 6093 Alu sequences and 1115 L1 sequences to mask 5' ESTs containing such repeat sequences from further consideration. 5' ESTs including THE and MER repeats, SSTR sequences or satellite, micro-satellite, or telomeric repeats were also eliminated from further consideration. On average, 11.5% of

-25-

the sequences in the libraries contained repeat sequences. Of this 11.5%, 7% contained Alu repeats, 3.3% contained L1 repeats and the remaining 1.2% were derived from the other types of repetitive sequences which were screened. These percentages are consistent with those found in cDNA libraries prepared by other groups. For example, the cDNA libraries of Adams et al. contained between 0% and 7.4% Alu repeats depending on the source of the RNA which was

5   used to prepare the cDNA library (Adams et al., *Nature* 377:174, 1996).

-     The sequences of those 5' ESTs remaining after the elimination of undesirable sequences were compared with the sequences of known human mRNAs to determine the accuracy of the sequencing procedures described above.

EXAMPLE 19

Measurement of Sequencing Accuracy by Comparison to Known Sequences

10   To further determine the accuracy of the sequencing procedure described above, the sequences of 5' ESTs derived from known sequences were identified and compared to the known sequences. First, a FASTA analysis with overhangs shorter than 5 bp on both ends was conducted on the 5' ESTs to identify those matching an entry in the public human mRNA database. The 6655 5' ESTs which matched a known human mRNA were then realigned with their cognate mRNA and dynamic programming was used to include substitutions, insertions, and deletions in the list of

15   "errors" which would be recognized. Errors occurring in the last 10 bases of the 5' EST sequences were ignored to avoid the inclusion of spurious cloning sites in the analysis of sequencing accuracy.

This analysis revealed that the sequences incorporated in the NETGENE™ database had an accuracy of more than 99.5%.

To determine the efficiency with which the above selection procedures select cDNAs which include the 5' ends

20   of their corresponding mRNAs, the following analysis was performed.

EXAMPLE 20

Determination of Efficiency of 5' EST Selection

To determine the efficiency at which the above selection procedures isolated 5' ESTs which included sequences close to the 5' end of the mRNAs from which they were derived, the sequences of the ends of the 5' ESTs

25   which were derived from the elongation factor 1 subunit α and ferritin heavy chain genes were compared to the known cDNA sequences for these genes. Since the transcription start sites for the elongation factor 1 subunit α and ferritin heavy chain are well characterized, they may be used to determine the percentage of 5' ESTs derived from these genes which included the authentic transcription start sites.

For both genes, more than 95% of the cDNAs included sequences close to or upstream of the 5' end of the

30   corresponding mRNAs.

To extend the analysis of the reliability of the procedures for isolating 5' ESTs from ESTs in the NETGENE™ database, a similar analysis was conducted using a database composed of human mRNA sequences extracted from GenBank database release 97 for comparison. For those 5' ESTs derived from mRNAs included in the GeneBank database, more than 85% had their 5' ends close to the 5' ends of the known sequence. As some of the mRNA

-26-

sequences available in the GenBank database are deduced from genomic sequences, a 5' end matching with these sequences will be counted as an internal match. Thus, the method used here underestimates the yield of ESTs including the authentic 5' ends of their corresponding mRNAs.

The EST libraries made above included multiple 5' ESTs derived from the same mRNA. The sequences of such 5' ESTs were compared to one another and the longest 5' ESTs for each mRNA were identified. Overlapping cDNAs were assembled into continuous sequences (contigs). The resulting continuous sequences were then compared to public databases to gauge their similarity to known sequences, as described in Example 21 below.

EXAMPLE 21

Clustering of the 5' ESTs and Calculation of Novelty Indices for cDNA Libraries

For each sequenced EST library, the sequences were clustered by the 5' end. Each sequence in the library was compared to the others with BLASTN2 (direct strand, parameters S = 107). ESTs with High Scoring Segment Pairs (HSPs) at least 25 bp long, having 95% identical bases and beginning closer than 10 bp from each EST 5' end were grouped. The longest sequence found in the cluster was used as representative of the cluster. A global clustering between libraries was then performed leading to the definition of super-contigs.

To assess the yield of new sequences within the EST libraries, a novelty rate (NR) was defined as: NR = 100 X (Number of new unique sequences found in the library/Total number of sequences from the library). Typically, novelty rating range between 10% and 41% depending on the tissue from which the EST library was obtained. For most of the libraries, the random sequencing of 5' EST libraries was pursued until the novelty rate reached 20%.

Following characterization as described above, the collection of 5' ESTs in NETGENE™ was screened to identify those 5' ESTs bearing potential signal sequences as described in Example 22 below.

EXAMPLE 22

Identification of Potential Signal Sequences in 5' ESTs

The 5' ESTs in the NETGENE™ database were screened to identify those having an uninterrupted open reading frame (ORF) longer than 45 nucleotides beginning with an ATG codon and extending to the end of the EST. Approximately half of the cDNA sequences in NETGENE™ contained such an ORF. The ORFs of these 5' ESTs were searched to identify potential signal motifs using slight modifications of the procedures disclosed in Von Heijne, G. A New Method for Predicting Signal Sequence Cleavage Sites. **Nucleic Acids Res.** 14:4683-4690 (1986). Those 5' EST sequences encoding a 15 amino acid long stretch with a score of at least 3.5 in the Von Heijne signal peptide identification matrix were considered to possess a signal sequence. Those 5' ESTs which matched a known human mRNA or EST sequence and had a 5' end more than 20 nucleotides downstream of the known 5' end were excluded from further analysis. The remaining cDNAs having signal sequences therein were included in a database called SIGNALTAG™.

To confirm the accuracy of the above method for identifying signal sequences, the analysis of Example 23 was performed.

·27·

EXAMPLE 23

Confirmation of Accuracy of Identification of Potential Signal Sequences in 5′ ESTs

The accuracy of the above procedure for identifying signal sequences encoding signal peptides was evaluated by applying the method to the 43 amino terminal amino acids of all human SwissProt proteins. The computed Von Heijne

5   score for each protein was compared with the known characterization of the protein as being a secreted protein or a non-secreted protein. In this manner, the number of non-secreted proteins having a score higher than 3.5 (false positives) and the number of secreted proteins having a score lower than 3.5 (false negatives) could be calculated.

Using the results of the above analysis, the probability that a peptide encoded by the 5′ region of the mRNA is in fact a genuine signal peptide based on its Von Heijne's score was calculated based on either the assumption that 10%

10   of human proteins are secreted or the assumption that 20% of human proteins are secreted. The results of this analysis are shown in Figures 2 and 3.

Using the above method of identifying secretory proteins, 5′ ESTs for human glucagon, gamma interferon induced monokine precursor, secreted cyclophilin-like protein, human pleiotropin, and human biotinidase precursor all of which are polypeptides which are known to be secreted, were obtained. Thus, the above method successfully identified

15   those 5′ ESTs which encode a signal peptide.

To confirm that the signal peptide encoded by the 5′ ESTs actually functions as a signal peptide, the signal sequences from the 5′ ESTs may be cloned into a vector designed for the identification of signal peptides. Some signal peptide identification vectors are designed to confer the ability to grow in selective medium on host cells which have a signal sequence operably inserted into the vector. For example, to confirm that a 5′ EST encodes a genuine signal

20   peptide, the signal sequence of the 5′ EST may be inserted upstream and in frame with a non-secreted form of the yeast invertase gene in signal peptide selection vectors such as those described in U.S. Patent No. 5,536,637. Growth of host cells containing signal sequence selection vectors having the signal sequence from the 5′ EST inserted therein confirms that the 5′ EST encodes a genuine signal peptide.

Alternatively, the presence of a signal peptide may be confirmed by cloning the extended cDNAs obtained using

25   the ESTs into expression vectors such as pXT1 (as described below), or by constructing promoter-signal sequence-reporter gene vectors which encode fusion proteins between the signal peptide and an assayable reporter protein. After introduction of these vectors into a suitable host cell, such as COS cells or NIH 3T3 cells, the growth medium may be harvested and analyzed for the presence of the secreted protein. The medium from these cells is compared to the medium from cells containing vectors lacking the signal sequence or extended cDNA insert to identify vectors which

30   encode a functional signal peptide or an authentic secreted protein.

Those 5′ ESTs which encoded a signal peptide, as determined by the method of Example 22 above, were further grouped into four categories based on their homology to known sequences. The categorization of the 5′ ESTs is described in Example 24 below.

EXAMPLE 24

-28-

<u>Categorization of 5' ESTs Encoding a Signal Peptide</u>

Those 5' ESTs having a sequence not matching any known vertebrate sequence nor any publicly available EST sequence were designated "new." Of the sequences in the SIGNALTAG™ database, 947 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

5   Those 5' ESTs having a sequence not matching any vertebrate sequence but matching a publicly known EST were designated "EST-ext", provided that the known EST sequence was extended by at least 40 nucleotides in the 5' direction. Of the sequences in the SIGNALTAG™ database, 150 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those ESTs not matching any vertebrate sequence but matching a publicly known EST without extending the

10   known EST by at least 40 nucleotides in the 5' direction were designated "EST." Of the sequences in the SIGNALTAG™ database, 599 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those 5' ESTs matching a human mRNA sequence but extending the known sequence by at least 40 nucleotides in the 5' direction were designated "VERT-ext." Of the sequences in the SIGNALTAG™ database, 23 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category. Included in this category was a 5' EST which

15   extended the known sequence of the human translocase mRNA by more than 200 bases in the 5' direction. A 5' EST which extended the sequence of a human tumor suppressor gene in the 5' direction was also identified.

Figure 4 shows the distribution of 5' ESTs in each category and the number of 5' ESTs in each category having a given minimum von Heijne's score.

Each of the 5' ESTs was categorized based on the tissue from which its corresponding mRNA was obtained,

20   as described below in Example 25.

EXAMPLE 25

<u>Categorization of Expression Patterns</u>

Figure 5 shows the tissues from which the mRNAs corresponding to the 5' ESTs in each of the above described categories were obtained.

25   In addition to categorizing the 5' ESTs by the tissue from which the cDNA library in which they were first identified was obtained, the spatial and temporal expression patterns of the mRNAs corresponding to the 5' ESTs, as well as their expression levels, may be determined as described in Example 26 below. Characterization of the spatial and temporal expression patterns and expression levels of these mRNAs is useful for constructing expression vectors capable of producing a desired level of gene product in a desired spatial or temporal manner, as will be discussed in more detail

30   below.

In addition, 5' ESTs whose corresponding mRNAs are associated with disease states may also be identified. For example, a particular disease may result from lack of expression, over expression, or under expression of an mRNA corresponding to a 5' EST. By comparing mRNA expression patterns and quantities in samples taken from healthy

·29·

individuals with those from individuals suffering from a particular disease, 5' ESTs responsible for the disease may be identified.

It will be appreciated that the results of the above characterization procedures for 5' ESTs also apply to extended cDNAs (obtainable as described below) which contain sequences adjacent to the 5' ESTs. It will also be

5   appreciated that if it is desired to defer characterization until extended cDNAs have been obtained rather than characterizing the ESTs themselves, the above characterization procedures can be applied to characterize the extended cDNAs after their isolation.

EXAMPLE 26

Evaluation of Expression Levels and Patterns of mRNAs

10   Corresponding to 5' ESTs or Extended cDNAs

Expression levels and patterns of mRNAs corresponding to 5' ESTs or extended cDNAs (obtainable as described below) may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277. Briefly, a 5' EST, extended cDNA, or fragment thereof corresponding to the gene encoding the mRNA to be characterized is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7

15   or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the 5' EST or extended cDNA has 100 or more nucleotides. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with

20   ribonucleases specific for single-stranded RNA (i.e. RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

The 5' ESTs, extended cDNAs, or fragments thereof may also be tagged with nucleotide sequences for the

25   serial analysis of gene expression (SAGE) as disclosed in UK Patent Application No. 2 305 241 A. In this method, cDNAs are prepared from a cell, tissue, organism or other source of nucleic acid for which it is desired to determine gene expression patterns. The resulting cDNAs are separated into two pools. The cDNAs in each pool are cleaved with a first restriction endonuclease, called an "anchoring enzyme," having a recognition site which is likely to be present at least once in most cDNAs. The fragments which contain the 5' or 3' most region of the cleaved cDNA are isolated by binding

30   to a capture medium such as streptavidin coated beads. A first oligonucleotide linker having a first sequence for hybridization of an amplification primer and an internal restriction site for a "tagging endonuclease" is ligated to the digested cDNAs in the first pool. Digestion with the second endonuclease produces short "tag" fragments from the cDNAs.

A second oligonucleotide having a second sequence for hybridization of an amplification primer and an internal restriction site is ligated to the digested cDNAs in the second pool. The cDNA fragments in the second pool are also digested with the "tagging endonuclease" to generate short "tag" fragments derived from the cDNAs in the second pool. The "tags" resulting from digestion of the first and second pools with the anchoring enzyme and the tagging
5   endonuclease are ligated to one another to produce "ditags." In some embodiments, the ditags are concatamerized to produce ligation products containing from 2 to 200 ditags. The tag sequences are then determined and compared to the sequences of the 5' ESTs or extended cDNAs to determine which 5' ESTs or extended cDNAs are expressed in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of the 5' ESTs or extended cDNAs in the cell, tissue, organism, or other source of nucleic acids is obtained.

10   Quantitative analysis of gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of full length cDNAs (i.e. extended cDNAs which include the coding sequence for the signal peptide, the coding sequence for the mature protein, and a stop codon), extended cDNAs, 5' ESTs or fragments of the full length cDNAs, extended cDNAs, or 5' ESTs of sufficient length to permit specific detection of gene expression. Preferably, the fragments are at least 15 nucleotides in length. More
15   preferably, the fragments are at least 100 nucleotides in length. More preferably, the fragments are more than 100 nucleotides in length. In some embodiments the fragments may be more than 500 nucleotides in length.

        For example, quantitative analysis of gene expression may be performed with full length cDNAs, extended cDNAs, 5' ESTs, or fragments thereof in a complementary DNA microarray as described by Schena et al. (*Science* 270:467-470, 1995; *Proc. Natl. Acad. Sci. U.S.A.* 93:10614-10619, 1996). Full length cDNAs, extended cDNAs, 5'
20   ESTs or fragments thereof are amplified by PCR and arrayed from 96-well microtiter plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

25   Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm^2 microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential expression measurements are
30   obtained by taking the average of the ratios of two independent hybridizations.

        Quantitative analysis of the expression of genes may also be performed with full length cDNAs, extended cDNAs, 5' ESTs, or fragments thereof in complementary DNA arrays as described by Pietu et al. (Genome Research 6:492-503, 1996). The full length cDNAs, extended cDNAs, 5' ESTs or fragments thereof are PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides.

After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis of the 5' ESTs or extended cDNAs can be done through high density
5   nucleotide arrays as described by Lockhart et al. (Nature Biotechnology 14: 1675-1680, 1996) and Sosnowsky et al. (Proc. Natl. Acad. Sci. 94:1119-1123, 1997). Oligonucleotides of 15-50 nucleotides corresponding to sequences of the 5' ESTs or extended cDNAs are synthesized directly on the chip (Lockhart et al., *supra*) or synthesized and then addressed to the chip (Sosnowski et al., *supra*). Preferably, the oligonucleotides are about 20 nucleotides in length.

cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are
10   synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart et al., *supra* and application of different electric fields (Sosnowsky et al., Proc. Natl. Acad. Sci. 94:1119-1123)., the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a
15   differential expression of the mRNA corresponding to the 5' EST or extended cDNA from which the oligonucleotide sequence has been designed.

III. Use of 5' ESTs to Clone Extended cDNAs and to Clone the Corresponding Genomic DNAs

Once 5' ESTs which include the 5' end of the corresponding mRNAs have been selected using the procedures described above, they can be utilized to isolate extended cDNAs which contain sequences adjacent to the 5' ESTs. The
20   extended cDNAs may include the entire coding sequence of the protein encoded by the corresponding mRNA, including the authentic translation start site, the signal sequence, and the sequence encoding the mature protein remaining after cleavage of the signal peptide. Such extended cDNAs are referred to herein as "full length cDNAs." Alternatively, the extended cDNAs may include only the sequence encoding the mature protein remaining after cleavage of the signal peptide, or only the sequence encoding the signal peptide.
25   Example 27 below describes a general method for obtaining extended cDNAs. Example 28 below describes the cloning and sequencing of several extended cDNAs, including extended cDNAs which include the entire coding sequence and authentic 5' end of the corresponding mRNA for several secreted proteins.

The methods of Examples 27, 28, and 29 can also be used to obtain extended cDNAs which encode less than the entire coding sequence of the secreted proteins encoded by the genes corresponding to the 5' ESTs. In some
30   embodiments, the extended cDNAs isolated using these methods encode at least 10 amino acids of one of the proteins encoded by the sequences of SEQ ID NOs: 40-140 and 242-377. In further embodiments, the extended cDNAs encode at least 20 amino acids of the proteins encoded by the sequences of SEQ ID NOs: 40-140 and 242-377. In further embodiments, the extended cDNAs encode at least 30 amino amino acids of the sequences of SEQ ID NOs: 40-140 and

·32·

242-377. In a preferred embodiment, the extended cDNAs encode a full length protein sequence, which includes the

protein coding sequences of SEQ ID NOs: 40-140 and 242-377.

EXAMPLE 27

General Method for Using 5' ESTs to Clone and Sequence Extended cDNAs

5   The following general method has been used to quickly and efficiently isolate extended cDNAs including

sequence adjacent to the sequences of the 5' ESTs used to obtain them. This method may be applied to obtain extended

cDNAs for any 5' EST in the NETGENE™ database, including those 5' ESTs encoding secreted proteins. The method is

summarized in Figure 6.

          1. Obtaining Extended cDNAs

10 a) First strand synthesis

          The method takes advantage of the known 5' sequence of the mRNA. A reverse transcription reaction is

conducted on purified mRNA with a poly 14dT primer containing a 49 nucleotide sequence at its 5' end allowing the

addition of a known sequence at the end of the cDNA which corresponds to the 3' end of the mRNA. For example, the

primer may have the following sequence: 5'-ATC GTT GAG ACT CGT ACC AGC AGA GTC ACG AGA GAG ACT ACA CGG

15 TAC TGG TTT TTT TTT TTT TTVN -3' (SEQ ID NO:14). Those skilled in the art will appreciate that other sequences

may also be added to the poly dT sequence and used to prime the first strand synthesis. Using this primer and a reverse

transcriptase such as the Superscript II (Gibco BRL) or Rnase H Minus M-MLV (Promega) enzyme, a reverse transcript

anchored at the 3' polyA site of the RNAs is generated.

          After removal of the mRNA hybridized to the first cDNA strand by alkaline hydrolysis, the products of the

20 alkaline hydrolysis and the residual poly dT primer are eliminated with an exclusion column such as an AcA34 (Biosepra)

matrix as explained in Example 11.

b) Second strand synthesis

          A pair of nested primers on each end is designed based on the known 5' sequence from the 5' EST and the

known 3' end added by the poly dT primer used in the first strand synthesis. Software used to design primers are either

25 based on GC content and melting temperatures of oligonucleotides, such as OSP (Illier and Green, *PCR Meth. Appl.*

1:124-128, 1991), or based on the octamer frequency disparity method (Griffais et al., *Nucleic Acids Res.* 19: 3887-

3891, 1991 such as PC-Rare (http://bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html).

          Preferably, the nested primers at the 5' end are separated from one another by four to nine bases. The 5'

primer sequences may be selected to have melting temperatures and specificities suitable for use in PCR.

30   Preferably, the nested primers at the 3' end are separated from one another by four to nine bases. For

example, the nested 3' primers may have the following sequences: 5' CCA GCA GAG TCA CGA GAG AGA CTA CAC GG

-3'(SEQ ID NO:15), and 5'- CAC GAG AGA GAC TAC ACG GTA CTG G 3' (SEQ ID NO:16). These primers were selected

because they have melting temperatures and specificities compatible with their use in PCR. However, those skilled in

the art will appreciate that other sequences may also be used as primers.

-33-

The first PCR run of 25 cycles is performed using the Advantage Tth Polymerase Mix (Clontech) and the outer

primer from each of the nested pairs. A second 20 cycle PCR using the same enzyme and the inner primer from each of

the nested pairs is then performed on 1/2500 of the first PCR product. Thereafter, the primers and nucleotides are

removed.

5   2. Sequencing of Full Length Extended cDNAs or Fragments Thereof

Due to the lack of position constraints on the design of 5' nested primers compatible for PCR use using the

OSP software, amplicons of two types are obtained. Preferably, the second 5' primer is located upstream of the

translation initiation codon thus yielding a nested PCR product containing the whole coding sequence. Such a full length

extended cDNA undergoes a direct cloning procedure as described in section a below. However, in some cases, the

10   second 5' primer is located downstream of the translation initiation codon, thereby yielding a PCR product containing

only part of the ORF. Such incomplete PCR products are submitted to a modified procedure described in section b

below.

a) Nested PCR products containing complete ORFs

When the resulting nested PCR product contains the complete coding sequence, as predicted from the 5'EST

15   sequence, it is cloned in an appropriate vector such as pED6dpc2, as described in section 3.

b) Nested PCR products containing incomplete ORFs

When the amplicon does not contain the complete coding sequence, intermediate steps are necessary to obtain

both the complete coding sequence and a PCR product containing the full coding sequence. The complete coding

sequence can be assembled from several partial sequences determined directly from different PCR products as described

20   in the following section.

Once the full coding sequence has been completely determined, new primers compatible for PCR use are

designed to obtain amplicons containing the whole coding region. However, in such cases, 3' primers compatible for PCR

use are located inside the 3' UTR of the corresponding mRNA, thus yielding amplicons which lack part of this region, i.e.

the polyA tract and sometimes the polyadenylation signal, as illustrated in figure 6. Such full length extended cDNAs are

25   then cloned into an appropriate vector as described in section 3.

c) Sequencing extended cDNAs

Sequencing of extended cDNAs is performed using a Die Terminator approach with the AmpliTaq DNA

polymerase FS kit available from Perkin Elmer.

In order to sequence PCR fragments, primer walking is performed using software such as OSP to choose

30   primers and automated computer software such as ASMG (Sutton et al., *Genome Science Technol.* 1: 9-19, 1995) to

construct contigs of walking sequences including the initial 5' tag using minimum overlaps of 32 nucleotides. Preferably,

primer walking is performed until the sequences of full length cDNAs are obtained.

Completion of the sequencing of a given extended cDNA fragment is assessed as follows. Since sequences

located after a polyA tract are difficult to determine precisely in the case of uncloned products, sequencing and primer

·34·

walking processes for PCR products are interrupted when a polyA tract is identified in extended cDNAs obtained as described in case b. The sequence length is compared to the size of the nested PCR product obtained as described above. Due to the limited accuracy of the determination of the PCR product size by gel electrophoresis, a sequence is considered complete if the size of the obtained sequence is at least 70 % the size of the first nested PCR product. If the
5   length of the sequence determined from the computer analysis is not at least 70% of the length of the nested PCR product, these PCR products are cloned and the sequence of the insertion is determined. When Northern blot data are available, the size of the mRNA detected for a given PCR product is used to finally assess that the sequence is complete. Sequences which do not fulfill the above criteria are discarded and will undergo a new isolation procedure.

Sequence data of all extended cDNAs are then transferred to a proprietary database, where quality controls
10 and validation steps are carried out as described in example 15.

3. Cloning of Full Length Extended cDNAs

The PCR product containing the full coding sequence is then cloned in an appropriate vector. For example, the extended cDNAs can be cloned into the expression vector pED6dpc2 (DiscoverEase, Genetics Institute, Cambridge, MA) as follows. The structure of pED6dpc2 is shown in Figure 7. pED6dpc2 vector DNA is prepared with blunt ends by
15 performing an EcoRI digestion followed by a fill in reaction. The blunt ended vector is dephosphorylated. After removal of PCR primers and ethanol precipitation, the PCR product containing the full coding sequence or the extended cDNA obtained as described above is phosphorylated with a kinase subsequently removed by phenol:Sevag extraction and precipitation. The double stranded extended cDNA is then ligated to the vector and the resulting expression plasmid introduced into appropriate host cells.

20   _   Since the PCR products obtained as described above are blunt ended molecules that can be cloned in either direction, the orientation of several clones for each PCR product is determined. Then, 4 to 10 clones are ordered in microtiter plates and subjected to a PCR reaction using a first primer located in the vector close to the cloning site and a second primer located in the portion of the extended cDNA corresponding to the 3' end of the mRNA. This second primer may be the antisense primer used in anchored PCR in the case of direct cloning (case a) or the antisense primer located
25 inside the 3'UTR in the case of indirect cloning (case b). Clones in which the start codon of the extended cDNA is operably linked to the promoter in the vector so as to permit expression of the protein encoded by the extended cDNA are conserved and sequenced. In addition to the ends of cDNA inserts, approximately 50 bp of vector DNA on each side of the cDNA insert are also sequenced.

The cloned PCR products are then entirely sequenced according to the aforementioned procedure. In this case,
30 contig assembly of long fragments is then performed on walking sequences that have already contigated for uncloned PCR products during primer walking. Sequencing of cloned amplicons is complete when the resulting contigs include the whole coding region as well as overlapping sequences with vector DNA on both ends.

4. Computer Analysis of Full Length Extended cDNA

-35-

Sequences of all full length extended cDNAs are then submitted to further analysis as described below and using the parameters found in Table II with the following modifications. For screening of miscellaneous subdivisions of Genbank, FASTA was used instead of BLASTN and 15 nucleotide of homology was the limit instead of 17. For Alu detection, BLASTN was used with the following parameters: $S = 72$; identity $= 70\%$; and length $= 40$ nucleotides.

5   Polyadenylation signal and polyA tail which were not search for the 5' ESTs were searched. For polyadenylation signal detection the signal (AATAAA) was searched with one permissible mismatch in the last ten nucleotides preceding the 5' end of the polyA. For the polyA, a stretch of 8 amino acids in the last 20 nucleotides of the sequence was searched with BLAST2N in the sense strand with the following parameters ($W = 6$, $S = 10$, $E = 1000$, and identity $= 90\%$). Finally, patented sequences and ORF homologies were searched using, respectively, BLASTN and BLASTP on GenSEQ

10 (Derwent's database of patented nucleotide sequences) and SWISSPROT for ORFs with the following parameters ($W = 8$ and $B = 10$). Before examining the extended full length cDNAs for sequences of interest, extended cDNAs which are not of interest are searched as follows.

a) Elimination of undesired sequences

Although 5'ESTs were checked to remove contaminant sequences as described in Example 18, a last verification was

15 carried out to identify extended cDNAs sequences derived from undesired sequences such as vector RNAs, transfer RNAs, ribosomal rRNAs, mitochondrial RNAs, prokaryotic RNAs and fungal RNAs using the FASTA and BLASTN programs on both strands of extended cDNAs as described below.

To identify the extended cDNAs encoding vector RNAs, extended cDNAs are compared to the known sequences of vector RNA using the FASTA program. Sequences of extended cDNAs with more than 90% homology over

20 stretches of 15 nucleotides are identified as vector RNA.

To identify the extended cDNAs encoding tRNAs, extended cDNA sequences were compared to the sequences of 1190 known tRNAs obtained from EMBL release 38, of which 100 were human. Sequences of extended cDNAs having more than 80% homology over 60 nucleotides using FASTA were identified as tRNA.

To identify the extended cDNAs encoding rRNAs, extended cDNA sequences were compared to the sequences

25 of 2497 known rRNAs obtained from EMBL release 38, of which 73 were human. Sequences of extended cDNAs having more than 80% homology over stretches longer than 40 nucleotides using BLASTN were identified as rRNAs.

To identify the extended cDNAs encoding mtRNAs, extended cDNA sequences were compared to the sequences of the two known mitochondrial genomes for which the entire genomic sequences are available and all sequences transcribed from these mitochondrial genomes including tRNAs, rRNAs, and mRNAs for a total of 38

30 sequences. Sequences of extended cDNAs having more than 80% homology over stretches longer than 40 nucleotides using BLASTN were identified as mtRNAs.

Sequences which might have resulted from other exogenous contaminants were identified by comparing extended cDNA sequences to release 105 of Genbank bacterial and fungal divisions. Sequences of extended cDNAs

-36-

having more than 90% homology over 40 nucleotides using BLASTN were identified as exogenous prokaryotic or fungal
contaminants.

In addition, extended cDNAs were searched for different repeat sequences, including Alu sequences, L1
sequences, THE and MER repeats, SSTR sequences or satellite, micro-satellite, or telomeric repeats. Sequences of
5   extended cDNAs with more than 70% homology over 40 nucleotide stretches using BLASTN were identified as repeat
sequences and masked in further identification procedures. In addition, clones showing extensive homology to repeats ,
i.e., matches of either more than 50 nucleotides if the homology was at least 75% or more than 40 nucleotides if the
homology was at least 85% or more than 30 nucleotides if the homology was at least 90%, were flagged.

b) Identification of structural features

10   Structural features, e.g. polyA tail and polyadenylation signal, of the sequences of full length extended cDNAs
are subsequently determined as follows.

A polyA tail is defined as a homopolymeric stretch of at least 11 A with at most one alternative base within it.
The polyA tail search is restricted to the last 20 nt of the sequence and limited to stretches of 11 consecutive A's
because sequencing reactions are often not readable after such a polyA stretch. Stretches with 100% homology over 6
15   nucleotides are identified as polyA tails.

To search for a polyadenylation signal, the polyA tail is clipped from the full-length sequence. The 50 bp
preceding the polyA tail are searched for the canonic polyadenylation AAUAAA signal allowing one mismatch to account
for possible sequencing errors and known variation in the canonical sequence of the polyadenylation signal.

c) Identification of functional features

20   _   Functional features, e.g. ORFs and signal sequences, of the sequences of full length extended cDNAs were
subsequently determined as follows.

The 3 upper strand frames of extended cDNAs are searched for ORFs defined as the maximum length
fragments beginning with a translation initiation codon and ending with a stop codon. ORFs encoding at least 20 amino
acids are preferred.

25   Each found ORF is then scanned for the presence of a signal peptide in the first 50 amino-acids or, where
appropriate, within shorter regions down to 20 amino acids or less in the ORF, using the matrix method of von Heijne
(Nuc. Acids Res. 14: 4683-4690 (1986)) and the modification described in Example 22.

d) Homology to either nucleotidic or proteic sequences

Sequences of full length extended cDNAs are then compared to known sequences on a nucleotidic or proteic
30   basis.

Sequences of full length extended cDNAs are compared to the following known nucleic acid sequences:
vertebrate sequences (Genbank) , EST sequences (Genbank), patented sequences (Geneseqn) and recently identified
sequences (Genbank daily releases) available at the time of filing for the priority documents. Full length cDNA sequences
are also compared to the sequences of a private database (Genset internal sequences) in order to find sequences that

have already been identified by applicants. Sequences of full length extended cDNAs with more than 90% homology over 30 nucleotides using either BLASTN or BLAST2N as indicated in Table III are identified as sequences that have already been described. Matching vertebrate sequences are subsequently examined using FASTA; full length extended cDNAs with more than 70% homology over 30 nucleotides are identified as sequences that have already been described.

5   ORFs encoded by full length extended cDNAs as defined in section c) are subsequently compared to known amino acid sequences found in Swissprot release CHP, PIR release PIR# and Genpept release GPEPT public databases using BLASTP with the parameter W = 8 and allowing a maximum of 10 matches. Sequences of full length extended cDNAs showing extensive homology to known protein sequences are recognized as already identified proteins.

In addition, the three-frame conceptual translation products of the top strand of full length extended cDNAs
10   are compared to publicly known amino acid sequences of Swissprot using BLASTX with the parameter E = 0.001. Sequences of full length extended cDNAs with more than 70% homology over 30 amino acid stretches are detected as already identified proteins.

5. Selection of Cloned Full Length Sequences of the Present Invention

Cloned full length extended cDNA sequences that have already been characterized by the aforementioned
15   computer analysis are then submitted to an automatic procedure in order to preselect full length extended cDNAs containing sequences of interest.

a) Automatic sequence preselection

All complete cloned full length extended cDNAs clipped for vector on both ends are considered. First, a negative selection is operated in order to eliminate unwanted cloned sequences resulting from either contaminants or
20   PCR artifacts as follows. Sequences matching contaminant sequences such as vector RNA, tRNA, mtRNA, rRNA sequences are discarded as well as those encoding ORF sequences exhibiting extensive homology to repeats as defined in section 4 a). Sequences obtained by direct cloning using nested primers on 5' and 3' tags (section 1. case a) but lacking polyA tail are discarded. Only ORFs containing a signal peptide and ending either before the polyA tail (case a) or before the end of the cloned 3'UTR (case b) are kept. Then, ORFs containing unlikely mature proteins such as mature
25   proteins which size is less than 20 amino acids or less than 25% of the immature protein size are eliminated.

In the selection of the OFR, priority was given to the ORF and the frame corresponding to the polypeptides described in SignalTag Patents (United States Patent Application Serial Nos: 08/905,223; 08/905,135; 08/905,051; 08/905,144; 08/905,279; 08/904,468; 08/905,134; and 08/905,133). If the ORF was not found among the OFRs described in the SignalTag Patents, the ORF encoding the signal peptide with the highest score according to Von Heijne
30   method as defined in Example 22 was chosen. If the scores were identical, then the longest ORF was chosen.

Sequences of full length extended cDNA clones are then compared pairwise with BLAST after masking of the repeat sequences. Sequences containing at least 90% homology over 30 nucleotides are clustered in the same class. Each cluster is then subjected to a cluster analysis that detects sequences resulting from internal priming or from

alternative splicing, identical sequences or sequences with several frameshifts. This automatic analysis serves as a basis for manual selection of the sequences.

b) Manual sequence selection

Manual selection is carried out using automatically generated reports for each sequenced full length extended
5   cDNA clone. During this manual procedures, a selection is operated between clones belonging to the same class as follows. ORF sequences encoded by clones belonging to the same class are aligned and compared. If the homology between nucleotidic sequences of clones belonging to the same class is more than 90% over 30 nucleotide stretches or if the homology between amino acid sequences of clones belonging to the same class is more than 80% over 20 amino acid stretches, than the clones are considered as being identical. The chosen ORF is the best one according to the
10 criteria mentioned below. If the nucleotide and amino acid homologies are less than 90% and 80% respectively, the clones are said to encode distinct proteins which can be both selected if they contain sequences of interest.

Selection of full length extended cDNA clones encoding sequences of interest is performed using the following criteria. Structural parameters (initial tag, polyadenylation site and signal) are first checked. Then, homologies with known nucleic acids and proteins are examined in order to determine whether the clone sequence match a known
15 nucleic/proteic sequence and, in the latter case, its covering rate and the date at which the sequence became public. If there is no extensive match with sequences other than ESTs or genomic DNA, or if the clone sequence brings substantial new information, such as encoding a protein resulting from alternative slicing of an mRNA coding for an already known protein, the sequence is kept. Examples of such cloned full length extended cDNAs containing sequences of interest are described in Example 28. Sequences resulting from chimera or double inserts as assessed by homology to other
20 sequences are discarded during this procedure.

EXAMPLE 28

Cloning and Sequencing of Extended cDNAs

The procedure described in Example 27 above was used to obtain the extended cDNAs of the present invention. Using this approach, the full length cDNA of SEQ ID NO:17 was obtained. This cDNA falls into the "EST-
25 ext" category described above and encodes the signal peptide MKKVLLLITAILAVAVG (SEQ ID NO: 18) having a von Heijne score of 8.2.

The full length cDNA of SEQ ID NO:19 was also obtained using this procedure. This cDNA falls into the "EST-ext" category described above and encodes the signal peptide MWWFQQGLSFLPSALVIWTSA (SEQ ID NO:20) having a von Heijne score of 5.5.

30   Another full length cDNA obtained using the procedure described above has the sequence of SEQ ID NO:21. This cDNA, falls into the "EST-ext" category described above and encodes the signal peptide MVLTTLPSANSANSPVNMPTTGPNSLSYASSALSPCLT (SEQ ID NO:22) having a von Heijne score of 5.9.

·39·

The above procedure was also used to obtain a full length cDNA having the sequence of SEQ ID NO:23. This cDNA falls into the "EST-ext" category described above and encodes the signal peptide ILSTVTALTFAXA (SEQ ID NO:24) having a von Heijne score of 5.5.

The full length cDNA of SEQ ID NO:25 was also obtained using this procedure. This cDNA falls into the "new" 5 category described above and encodes a signal peptide LVLTLCTLPLAVA (SEQ ID NO:26) having a von Heijne score of 10.1.

The full length cDNA of SEQ ID NO:27 was also obtained using this procedure. This cDNA falls into the "new" category described above and encodes a signal peptide LWLLFFLVTAIHA (SEQ ID NO:28) having a von Heijne score of 10.7.

10 The above procedures were also used to obtain the extended cDNAs of the present invention. 5' ESTs expressed in a variety of tissues were obtained as described above. The appended sequence listing provides the tissues from which the extended cDNAs were obtained. It will be appreciated that the extended cDNAs may also be expressed in tissues other than the tissue listed in the sequence listing.

5' ESTs obtained as described above were used to obtain extended cDNAs having the sequences of SEQ ID 15 NOs: 40-140 and 242-377. Table IV provides the sequence identification numbers of the extended cDNAs of the present invention, the locations of the full coding sequences in SEQ ID NOs: 40-140 and 242-377 (i.e. the nucleotides encoding both the signal peptide and the mature protein, listed under the heading FCS location in Table IV), the locations of the nucleotides in SEQ ID NOs: 40-140 and 242-377 which encode the signal peptides (listed under the heading SigPep Location in Table IV), the locations of the nucleotides in SEQ ID NOs: 40-140 and 242-377 which encode the mature 20 proteins generated by cleavage of the signal peptides (listed under the heading Mature Polypeptide Location in Table IV), the locations in SEQ ID NOs: 40-140 and 242-377 of stop codons (listed under the heading Stop Codon Location in Table IV), the locations in SEQ ID NOs: 40-140 and 242-377 of polyA signals (listed under the heading Poly A Signal Location in Table IV) and the locations of polyA sites (listed under the heading Poly A Site Location in Table IV).

The polypeptides encoded by the extended cDNAs were screened for the presence of known structural or 25 functional motifs or for the presence of signatures, small amino acid sequences which are well conserved amongst the members of a protein family. The conserved regions have been used to derive consensus patterns or matrices included in the PROSITE data bank, in particular in the file prosite.dat (Release 13.0 of November 1995, located at http://expasy.hcuge.ch/sprot/prosite.html. Prosite_convert and prosite_scan programs (http://ulrec3.unil.ch/ftpserveur/prosite_scan) were used to find signatures on the extended cDNAs.

30 For each pattern obtained with the prosite_convert program from the prosite.dat file, the accuracy of the detection on a new protein sequence has been tested by evaluating the frequency of irrelevant hits on the population of human secreted proteins included in the data bank SWISSPROT. The ratio between the number of hits on shuffled proteins (with a window size of 20 amino acids) and the number of hits on native (unshuffled) proteins was used as an index. Every pattern for which the ration was greater than 20% (one hit on shuffled proteins for 5 hits on native

proteins) was skipped during the search with prosite_scan. The program used to shuffle protein sequences (db_shuffled) and the program used to determine the statistics for each pattern in the protein data banks (prosite_statistics) are available on the ftp site http://ulrec3.unil.ch/ftpserveur/prosite_scan.

Table V lists the sequence identification numbers of the polypeptides of SEQ ID NOs: 141-241 and 378-513,
5   the locations of the amino acid residues of SEQ ID NOs: 141-241 and 378-513 in the full length polypeptide (second column), the locations of the amino acid residues of SEQ ID NOs: 141-241 and 378-513 in the signal peptides (third column), and the locations of the amino acid residues of SEQ ID NOs: 141-241 and 378-513 in the mature polypeptide created by cleaving the signal peptide from the full length polypeptide (fourth column).

The nucleotide sequences of the sequences of SEQ ID NOs: 40-140 and 242-377 and the amino acid sequences
10   encoded by SEQ ID NOs: 40-140 and 242-377 (i.e. amino acid sequences of SEQ ID NOs: 141-241 and 378-513) are provided in the appended sequence listing. In some instances, the sequences are preliminary and may include some incorrect or ambiguous sequences or amino acids. The sequences of SEQ ID NOs: 40-140 and 242-377 can readily be screened for any errors therein and any sequence ambiguities can be resolved by resequencing a fragment containing such errors or ambiguities on both strands. Nucleic acid fragments for resolving sequencing errors or ambiguities may be
15   obtained from the deposited clones or can be isolated using the techniques described herein. Resolution of any such ambiguities or errors may be facilitated by using primers which hybridize to sequences located close to the ambiguous or erroneous sequences. For example, the primers may hybridize to sequences within 50-75 bases of the ambiguity or error. Upon resolution of an error or ambiguity, the corresponding corrections can be made in the protein sequences encoded by the DNA containing the error or ambiguity. For example, in the sequences of the present invention, ambiguities
20   in the sequence of SEQ ID NO: 131 were resolved. The amino acid sequence of the protein encoded by a particular clone can also be determined by expression of the clone in a suitable host cell, collecting the protein, and determining its sequence.

For each amino acid sequence, Applicants have identified-what they have determined to be the reading frame best identifiable with sequence information available at the time of filing. Some of the amino acid sequences may
25   contain "Xaa" designators. These "Xaa" designators indicate either (1) a residue which cannot be identified because of nucleotide sequence ambiguity or (2) a stop codon in the determined sequence where Applicants believe one should not exist (if the sequence were determined more accurately).

Cells containing the extended cDNAs (SEQ ID NOs: 40-140 and 242-377) of the present invention in the vector pED6dpc2, are maintained in permanent deposit by the inventors at Genset, S.A., 24 Rue Royale, 75008 Paris, France.
30   Pools of cells containing the extended cDNAs (SEQ ID NOs: 40-140 and 242-377), from which cells containing a particular polynucleotide are obtainable, were deposited with the American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209 or the European Collection of Cell Cultures, Vaccine Research and Production Laboratory, Public Health Laboratory Service, Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom. Each extended cDNA clone has been transfected into separate bacterial cells (E-

-41-

coli) for this composite deposit. Table VI lists the deposit numbers of the clones containing the extended cDNAs of the present invention. Table VII provides the internal designation number assigned to each SEQ ID NO and indicates whether the sequence is a nucleic acid sequence or a protein sequence.

Each extended cDNA can be removed from the pED6dpc2 vector in which it was deposited by performing a
5   NotI, PstI double digestion to produce the appropriate fragment for each clone. The proteins encoded by the extended cDNAs may also be expressed from the promoter in pED6dpc2.

Bacterial cells containing a particular clone can be obtained from the composite deposit as follows:

An oligonucleotide probe or probes should be designed to the sequence that is known for that particular clone. This sequence can be derived from the sequences provided herein, or from a combination of those sequences. The design
10   of the oligonucleotide probe should preferably follow these parameters:

(a) It should be designed to an area of the sequence which has the fewest ambiguous bases ("N's"), if any;

(b) Preferably, the probe is designed tc have a T_m of approx. 80°C (assuming 2 degrees for each A or T and 4 degrees for each G or C). However, probes having melting temperatures between 40 °C and 80 °C may also be used provided that specificity is not lost.

15   The oligonucleotide should preferably be labeled with ($\cdot$[^{32}P]ATP (specific activity 6000 Ci/mmole) and T4 polynucleotide kinase using commonly employed techniques for labeling oligonucleotides. Other labeling techniques can also be used. Unincorporated label should preferably be removed by gel filtration chromatography or other established methods. The amount of radioactivity incorporated into the probe should be quantified by measurement in a scintillation counter. Preferably, specific activity of the resulting probe should be approximately 4×10^6 dpm/pmole.

20   The bacterial culture containing the pool of full-length clones should preferably be thawed and 100 μl of the stock used to inoculate a sterile culture flask containing 25 ml of sterile L-broth containing ampicillin at 100 ug/ml. The culture should preferably be grown to saturation at 37°C, and the saturated culture should preferably be diluted in fresh L-broth. Aliquots of these dilutions should preferably be plated to determine the dilution and volume which will yield approximately 5000 distinct and well-separated colonies on solid bacteriological media containing L-broth containing
25   ampicillin at 100 μg/ml and agar at 1.5% in a 150 mm petri dish when grown overnight at 37°C. Other known methods of obtaining distinct, well-separated colonies can also be employed.

Standard colony hybridization procedures should then be used to transfer the colonies to nitrocellulose filters and lyse, denature and bake them.

The filter is then preferably incubated at 65°C for 1 hour with gentle agitation in 6X SSC (20X stock is
30   175.3 g NaC1/liter, 88.2 g Na citrate/liter, adjusted to pH 7.0 with NaOH) containing 0.5% SDS, 100 pg/ml of yeast RNA, and 10 mM EDTA (approximately 10 mL per 150 mm filter). Preferably, the probe is then added to the hybridization mix at a concentration greater than or equal to 1×10^6 dpm/mL. The filter is then preferably incubated at 65°C with gentle agitation overnight. The filter is then preferably washed in 500 mL of 2X SSC/0.1% SDS at room temperature with gentle shaking for 15 minutes. A third wash with 0.1X SSC/0.5% SDS at 65°C for 30 minutes to

-42-

1 hour is optional. The filter is then preferably dried and subjected to autoradiography for sufficient time to visualize the positives on the X-ray film. Other known hybridization methods can also be employed.

The positive colonies are picked, grown in culture, and plasmid DNA isolated using standard procedures. The clones can then be verified by restriction analysis, hybridization analysis, or DNA sequencing.

5   The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the extended cDNA insertion. For example, a PCR reaction may be conducted using a primer having the sequence GGCCATACACTTGAGTGAC (SEQ ID NO:38) and a primer having the sequence ATATAGACAAACGCACACC (SEQ. ID. NO:39). The PCR product which corresponds to the extended cDNA can then be manipulated using standard cloning 10   techniques familiar to those skilled in the art.

In addition to PCR based methods for obtaining extended cDNAs, traditional hybridization based methods may also be employed. These methods may also be used to obtain the genomic DNAs which encode the mRNAs from which the 5' ESTs were derived, mRNAs corresponding to the extended cDNAs, or nucleic acids which are homologous to extended cDNAs or 5' ESTs. Example 29 below provides an example of such methods.

15   **EXAMPLE 29**

                        Methods for Obtaining Extended cDNAs or Nucleic

                        Acids Homologous to Extended cDNAs or 5' ESTs

A full length cDNA library can be made using the strategies described in Examples 13, 14, 15, and 16 above by replacing the random nonamer used in Example 14 with an oligo-dT primer. For instance, the oligonucleotide of SEQ ID 20   NO:14 may be used.

Alternatively, a cDNA library or genomic DNA library may be obtained from a commercial source or made using techniques familiar to those skilled in the art. The library includes cDNAs which are derived from the mRNA corresponding to a 5' EST or which have homology to an extended cDNA or 5' EST. The cDNA library or genomic DNA library is hybridized to a detectable probe comprising at least 10 consecutive nucleotides from the 5' EST or extended 25   cDNA using conventional techniques. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST or extended cDNA. More preferably, the probe comprises at least 20-30 consecutive nucleotides from the 5' EST or extended cDNA. In some embodiments, the probe comprises at least 30 nucleotides from the 5' EST or extended cDNA. In other embodiments, the probe comprises at least 40, at least 50, at least 75, at least 100, at least 150, or at least 200 consecutive nucleotides from the 5' EST or extended cDNA.

30   Techniques for identifying cDNA clones in a cDNA library which hybridize to a given probe sequence are disclosed in Sambrook et al., Molecular Cloning: A Laboratory Manual 2d Ed., Cold Spring Harbor Laboratory Press, 1989. The same techniques may be used to isolate genomic DNAs.

Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are identified and isolated for further manipulation as follows. A probe comprising at least 10 consecutive nucleotides from the 5' EST or extended

-43-

cDNA is labeled with a detectable label such as a radioisotope or a fluorescent molecule. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5′ EST or extended cDNA. More preferably, the probe comprises 20-30 consecutive nucleotides from the 5′ EST or extended cDNA. In some embodiments, the probe comprises more than 30 nucleotides from the 5′ EST or extended cDNA. In some embodiments, the probe comprises at

5 least 40, at least 50, at least 75, at least 100, at least 150, or at least 200 consecutive nucleotides from the 5′ EST or extended cDNA.

Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, in vitro transcription, and non-radioactive techniques. The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter and denatured. After incubation of the filter with a blocking solution, the

10 filter is contacted with the labeled probe and incubated for a sufficient amount of time for the probe to hybridize to cDNAs or genomic DNAs containing a sequence capable of hybridizing to the probe.

By varying the stringency of the hybridization conditions used to identify extended cDNAs or genomic DNAs which hybridize to the detectable probe, extended cDNAS having different levels of homology to the probe can be identified and isolated. To identify extended cDNAs or genomic DNAs having a high degree of homology to the probe

15 sequence, the melting temperature of the probe may be calculated using the following formulas:

For probes between 14 and 70 nucleotides in length the melting temperature (Tm) is calculated using the formula: $Tm = 81.5 + 16.6(\log [Na+]) + 0.41(\text{fraction } G+C) - (600/N)$ where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation $Tm = 81.5 + 16.6(\log [Na+]) + 0.41(\text{fraction } G+C) - (0.63\% \text{ formamide}) - (600/N)$ where N is

20 the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100μg denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100μg denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook et al., supra.

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where

25 the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to extended cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25°C below the Tm. For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 15-25°C below the Tm. Preferably, for hybridizations in

30 6X SSC, the hybridization is conducted at approximately 68°C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

All of the foregoing hybridizations would be considered to be under "stringent" conditions. Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed

-44-

with 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour. Thereafter, the solution is washed at the hybridization temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

Extended cDNAs, nucleic acids homologous to extended cDNAs or 5' ESTs, or genomic DNAs which have hybridized to the probe are identified by autoradiography or other conventional techniques.

5   The above procedure may be modified to identify extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs having decreasing levels of homology to the probe sequence. For example, to obtain extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs of decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be decreased in increments of 5°C from 68°C to 42°C in a hybridization buffer having a Na+ concentration of approximately 1M. Following

10   hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C.

Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following

15   hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide.

Extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs which have hybridized to the probe are identified by autoradiography.

If it is desired to obtain nucleic acids homologous to extended cDNAs, such as allelic variants thereof or nucleic

20   acids encoding proteins related to the proteins encoded by the extended cDNAs, the level of homology between the hybridized nucleic acid and the extended cDNA or 5' EST used as the probe may readily be determined. To determine the level of homology between the hybridized nucleic acid and the extended cDNA or 5'EST from which the probe was derived, the nucleotide sequences of the hybridized nucleic acid and the extended cDNA or 5'EST from which the probe was derived are compared. For example, using the above methods, nucleic acids having at least 95% nucleic acid

25   homology to the extended cDNA or 5'EST from which the probe was derived may be obtained and identified. Similarly, by using progressively less stringent hybridization conditions one can obtain and identify nucleic acids having at least 90%, at least 85%, at least 80% or at least 75% homology to the extended cDNA or 5'EST from which the probe was derived. The level of homology between the hybridized nucleic acid and the extended cDNA or 5' EST used as the probe may be further determined using BLAST2N; parameters may be adapted depending on the sequence length and degree of

30   homology studied. In such comparisons, the default parameters or the parameters listed in Tables II and III may be used.

To determine whether a clone encodes a protein having a given amount of homology to the protein encoded by the extended cDNA or 5' EST, the amino acid sequence encoded by the extended cDNA or 5' EST is compared to the amino acid sequence encoded by the hybridizing nucleic acid. Homology is determined to exist when an amino acid sequence in the extended cDNA or 5' EST is closely related to an amino acid sequence in the hybridizing nucleic acid. A

sequence is closely related when it is identical to that of the extended cDNA or 5' EST or when it contains one or more amino acid substitutions therein in which amino acids having similar characteristics have been substituted for one another. Using the above methods, one can obtain nucleic acids encoding proteins having at least 95%, at least 90%, at least 85%, at least 80% or at least 75% homology to the proteins encoded by the extended cDNA or 5'EST from which

5 the probe was derived. Using the above methods and algorithms such as FASTA with parameters depending on the sequence length and degree of homology studied the level of homology may be determined. In determining the level of homology using FASTA, the default parameters or the parameters listed in Tables II or III may be used.

Alternatively, extended cDNAs may be prepared by obtaining mRNA from the tissue, cell, or organism of interest using mRNA preparation procedures utilizing poly A selection procedures or other techniques known to those

10 skilled in the art. A first primer capable of hybridizing to the poly A tail of the mRNA is hybridized to the mRNA and a reverse transcription reaction is performed to generate a first cDNA strand.

The first cDNA strand is hybridized to a second primer containing at least 10 consecutive nucleotides of the sequences of the 5' EST for which an extended cDNA is desired. Preferably, the primer comprises at least 12, 15, or 17 consecutive nucleotides from the sequences of the 5' EST. More preferably, the primer comprises 20-30 consecutive

15 nucleotides from the sequences of the 5' EST. In some embodiments, the primer comprises more than 30 nucleotides from the sequences of the 5' EST. If it is desired to obtain extended cDNAs containing the full protein coding sequence, including the authentic translation initiation site, the second primer used contains sequences located upstream of the translation initiation site. The second primer is extended to generate a second cDNA strand complementary to the first cDNA strand. Alternatively, RTPCR may be performed as described above using primers from both ends of the cDNA to

20 be obtained.

Extended cDNAs containing 5' fragments of the mRNA may be prepared by contacting an mRNA comprising the sequence of the 5' EST for which an extended cDNA is desired with a primer comprising at least 10 consecutive nucleotides of the sequences complementary to the 5' EST, hybridizing the primer to the mRNAs, and reverse transcribing the hybridized primer to make a first cDNA strand from the mRNAs. Preferably, the primer comprises at

25 least 12, 15, or 17 consecutive nucleotides from the 5' EST. More preferably, the primer comprises 20-30 consecutive nucleotides from the 5' EST.

Thereafter, a second cDNA strand complementary to the first cDNA strand is synthesized. The second cDNA strand may be made by hybridizing a primer complementary to sequences in the first cDNA strand to the first cDNA strand and extending the primer to generate the second cDNA strand.

30   The double stranded extended cDNAs made using the methods described above are isolated and cloned. The extended cDNAs may be cloned into vectors such as plasmids or viral vectors capable of replicating in an appropriate host cell. For example, the host cell may be a bacterial, mammalian, avian, or insect cell.

Techniques for isolating mRNA, reverse transcribing a primer hybridized to mRNA to generate a first cDNA strand, extending a primer to make a second cDNA strand complementary to the first cDNA strand, isolating the double

-46-

stranded cDNA and cloning the double stranded cDNA are well known to those skilled in the art and are described in Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997 and Sambrook et al. Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, 1989.

Alternatively, kits for obtaining full length cDNAs, such as the GeneTrapper (Cat. No. 10356-020, Gibco, BRL),
5 may be used for obtaining full length cDNAs or extended cDNAs. In this approach, full length or extended cDNAs are prepared from mRNA and cloned into double stranded phagemids. The cDNA library in the double stranded phagemids is then rendered single stranded by treatment with an endonuclease, such as the Gene II product of the phage F1, and Exonuclease III as described in the manual accompanying the GeneTrapper kit. A biotinylated oligonucleotide comprising the sequence of a 5' EST, or a fragment containing at least 10 nucleotides thereof, is hybridized to the single stranded
10 phagemids. Preferably, the fragment comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST. More preferably, the fragment comprises 20-30 consecutive nucleotides from the 5' EST. In some procedures, the fragment may comprise more than 30 consecutive nucleotides from the 5' EST. For example, the fragment may comprises at least 40, at least 50, at least 75, at least 100, at least 150, or at least 200 consecutive nucleotides from the 5' EST.

Hybrids between the biotinylated oligonucleotide and phagemids having inserts containing the 5' EST sequence
15 are isolated by incubating the hybrids with streptavidin coated paramagnetic beads and retrieving the beads with a magnet. Thereafter, the resulting phagemids containing the 5' EST sequence are released from the beads and converted into double stranded DNA using a primer specific for the 5' EST sequence. The resulting double stranded DNA is transformed into bacteria. Extended cDNAs containing the 5' EST sequence are identified by colony PCR or colony hybridization.

20   A plurality of extended cDNAs containing full length protein coding sequences or sequences encoding only the mature protein remaining after the signal peptide is cleaved may be provided as cDNA libraries for subsequent evaluation of the encoded proteins or use in diagnostic assays as described below.

IV. Expression of Proteins Encoded by Extended cDNAs Isolated Using 5' ESTs

Extended cDNAs containing the full protein coding sequences of their corresponding mRNAs or portions
25 thereof, such as cDNAs encoding the mature protein, may be used to express the secreted proteins or portions thereof which they encode as described in Example 30 below. If desired, the extended cDNAs may contain the sequences encoding the signal peptide to facilitate secretion of the expressed protein. It will be appreciated that a plurality of extended cDNAs containing the full protein coding sequences or portions thereof may be simultaneously cloned into expression vectors to create an expression library for analysis of the encoded proteins as described below.

30   EXAMPLE 30

Expression of the Proteins Encoded by Extended cDNAs or Portions Thereof

To express the proteins encoded by the extended cDNAs or portions thereof, nucleic acids containing the coding sequence for the proteins or portions thereof to be expressed are obtained as described in Examples 27-29 and cloned into a suitable expression vector. If desired, the nucleic acids may contain the sequences encoding the signal

peptide to facilitate secretion of the expressed protein. For example, the nucleic acid may comprise the sequence of one of SEQ ID NOs: 40-140 and 242-377 listed in Table IV and in the accompanying sequence listing. Alternatively, the nucleic acid may comprise those nucleotides which make up the full coding sequence of one of the sequences of SEQ ID NOs: 40-140 and 242-377 as defined in Table IV above.

5   It will be appreciated that should the extent of the full coding sequence (i.e. the sequence encoding the signal peptide and the mature protein resulting from cleavage of the signal peptide) differ from that listed in Table IV as a result of a sequencing error, reverse transcription or amplification error, mRNA splicing, post-translational modification of the encoded protein, enzymatic cleavage of the encoded protein, or other biological factors, one skilled in the art would be readily able to identify the extent of the full coding sequences in the sequences of SEQ ID NOs. 40-140 and 242-377.

10 For example, the sequence of SEQ ID NO: 115 represents an alternatively spliced transcript of a previously identified mRNA.. Accordingly, the scope of any claims herein relating to nucleic acids containing the full coding sequence of one of SEQ ID NOs. 40-140 and 242-377 is not to be construed as excluding any readily identifiable variations from or equivalents to the full coding sequences listed in Table IV Similarly, should the extent of the full length polypeptides differ from those indicated in Table V as a result of any of the preceding factors, the scope of claims relating to polypeptides

15 comprising the amino acid sequence of the full length polypeptides is not to be construed as excluding any readily identifiable variations from or equivalents to the sequences listed in Table V.

      Alternatively, the nucleic acid used to express the protein or portion thereof may comprise those nucleotides which encode the mature protein (i.e. the protein created by cleaving the signal peptide off) encoded by one of the sequences of SEQ ID NOs: 40-140 and 242-377 as defined in Table IV above.

20   It will be appreciated that should the extent of the sequence encoding the mature protein differ from that listed in Table IV as a result of a sequencing error, reverse transcription or amplification error, mRNA splicing, post-translational modification of the encoded protein, enzymatic cleavage of the encoded protein, or other biological factors, one skilled in the art would be readily able to identify the extent of the sequence encoding the mature protein in the sequences of SEQ ID NOs. 40-140 and 242-377. Accordingly, the scope of any claims herein relating to nucleic acids

25 containing the sequence encoding the mature protein encoded by one of SEQ ID Nos. 40-140 and 242-377 is not to be construed as excluding any readily identifiable variations from or equivalents to the sequences listed in Table IV. Thus, claims relating to nucleic acids containing the sequence encoding the mature protein encompass equivalents to the sequences listed in Table IV, such as sequences encoding biologically active proteins resulting from post-translational modification, enzymatic cleavage, or other readily identifiable variations from or equivalents to the secreted proteins in

30 addition to cleavage of the signal peptide. Similarly, should the extent of the mature polypeptides differ from those indicated in Table V as a result of any of the preceding factors, the scope of claims relating to polypeptides comprising the sequence of a mature protein included in the sequence of one of SEQ ID NOs. 141-241 and 378-513 is not to be construed as excluding any readily identifiable variations from or equivalents to the sequences listed in Table V. Thus, claims relating to polypeptides comprising the sequence of the mature protein encompass equivalents to the sequences

-48-

listed in Table IV, such as biologically active proteins resulting from post-translational modification, enzymatic cleavage, or other readily identifiable variations from or equivalents to the secreted proteins in addition to cleavage of the signal peptide. It will also be appreciated that should the biologically active form of the polypeptides included in the sequence of one of SEQ ID NOs. 141-241 and 378-513 or the nucleic acids encoding the biologically active form of the

5   polypeptides differ from those identified as the mature polypeptide in Table V or the nucleotides encoding the mature polypeptide in Table IV as a result of a sequencing error, reverse transcription or amplification error, mRNA splicing, post-translational modification of the encoded protein, enzymatic cleavage of the encoded protein, or other biological factors, one skilled in the art would be readily able to identify the amino acids in the biologically active form of the polypeptides and the nucleic acids encoding the biologically active form of the polypeptides. In such instances, the

10   claims relating to polypetides comprising the mature protein included in one of SEQ ID NOs. 141-241 and 378-513 or nucleic acids comprising the nucleotides of one of SEQ ID NOs. 40-140 and 242-377 encoding the mature protein shall not be construed to exclude any readily identifiable variations from the sequences listed in Table IV and Table V.

In some embodiments, the nucleic acid used to express the protein or portion thereof may comprise those nucleotides which encode the signal peptide encoded by one of the sequences of SEQ ID NOs: 40-140 and 242-377 as

15   defined in Table IV above.

It will be appreciated that should the extent of the sequence encoding the signal peptide differ from that listed in Table IV as a result of a sequencing error, reverse transcription or amplification error, mRNA splicing, post-translational modification of the encoded protein, enzymatic cleavage of the encoded protein, or other biological factors, one skilled in the art would be readily able to identify the extent of the sequence encoding the signal peptide in the

20   sequences of SEQ ID NOs. 40-140 and 242-377. Accordingly, the scope of any claims herein relating to nucleic acids containing the sequence encoding the signal peptide encoded by one of SEQ ID Nos. 40-140 and 242-377 is not to be construed as excluding any readily identifiable variations from the sequences listed in Table IV. Similarly, should the extent of the signal peptides differ from those indicated in Table V as a result of any of the preceding factors, the scope of claims relating to polypeptides comprising the sequence of a signal peptide included in the sequence of one of SEQ ID

25   NOs. 141-241 and 378-513 is not to be construed as excluding any readily identifiable variations from the sequences listed in Table V.

Alternatively, the nucleic acid may encode a polypeptide comprising at least 10 consecutive amino acids of one of the sequences of SEQ ID NOs: 141-241 and 378-513. In some embodiments, the nucleic acid may encode a polypeptide comprising at least 15 consecutive amino acids of one of the sequences of SEQ ID NOs: 141-241 and 378-

30   513. In other embodiments, the nucleic acid may encode a polypeptide comprising at least 25 consecutive amino acids of one of the sequences of SEQ ID NOs: 141-241 and 378-513. In other embodiments, the nucleic acid may encode a polypeptide comprising at least 60, at least 75, at least 100 or more than 100 consecutive amino acids of one of the sequences of SEQ ID Nos: 141-241 and 378-513.

The nucleic acids inserted into the expression vectors may also contain sequences upstream of the sequences encoding the signal peptide, such as sequences which regulate expression levels or sequences which confer tissue specific expression.

The nucleic acid encoding the protein or polypeptide to be expressed is operably linked to a promoter in an

5   expression vector using conventional cloning technology. The expression vector may be any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular

10 expression organism in which the expression vector is introduced, as explained by Hatfield, et al., U.S. Patent No. 5,082,767.

The following is provided as one exemplary method to express the proteins encoded by the extended cDNAs corresponding to the 5' ESTs or the nucleic acids described above. First, the methionine initiation codon for the gene and the poly A signal of the gene are identified. If the nucleic acid encoding the polypeptide to be expressed lacks a

15 methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the extended cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the *gag* gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct

20 allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The extended cDNA or portion thereof encoding the polypeptide to be expressed is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the extended cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5'primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the extended cDNA is positioned in frame with the poly A

25 signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and digested with BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri). Preferably the expressed protein is

30 released into the culture medium, thereby facilitating purification.

Alternatively, the extended cDNAs may be cloned into pED6dpc2 as described above. The resulting pED6dpc2 constructs may be transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded. Preferably, the protein expressed from the extended cDNA is released into the culture medium thereby facilitating purification.

·50·

Proteins in the culture medium are separated by gel electrophoresis. If desired, the proteins may be ammonium sulfate precipitated or separated based on size or charge prior to electrophoresis.

As a control, the expression vector lacking a cDNA insert is introduced into host cells or organisms and the proteins in the medium are harvested. The secreted proteins present in the medium are detected using techniques such

5   as Coomassie or silver staining or using antibodies against the protein encoded by the extended cDNA. Coomassie and silver staining techniques are familiar to those skilled in the art.

Antibodies capable of specifically recognizing the protein of interest may be generated using synthetic 15-mer peptides having a sequence encoded by the appropriate 5' EST, extended cDNA, or portion thereof. The synthetic peptides are injected into mice to generate antibody to the polypeptide encoded by the 5' EST, extended cDNA, or

10   portion thereof.

Secreted proteins from the host cells or organisms containing an expression vector which contains the extended cDNA derived from a 5' EST or a portion thereof are compared to those from the control cells or organism. The presence of a band in the medium from the cells containing the expression vector which is absent in the medium from the control cells indicates that the extended cDNA encodes a secreted protein. Generally, the band corresponding

15   to the protein encoded by the extended cDNA will have a mobility near that expected based on the number of amino acids in the open reading frame of the extended cDNA. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Alternatively, if the protein expressed from the above expression vectors does not contain sequences directing its secretion, the proteins expressed from host cells containing an expression vector containing an insert encoding a

20   secreted protein or portion thereof can be compared to the proteins expressed in host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the desired protein or portion thereof is being expressed. Generally, the band will have the mobility expected for the secreted protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such

25   as glycosylation, ubiquitination, or enzymatic cleavage.

The protein encoded by the extended cDNA may be purified using standard immunochromatography techniques. In such procedures, a solution containing the secreted protein, such as the culture medium or a cell extract, is applied to a column having antibodies against the secreted protein attached to the chromatography matrix. The secreted protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound

30   proteins. The specifically bound secreted protein is then released from the column and recovered using standard techniques.

If antibody production is not possible, the extended cDNA sequence or portion thereof may be incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies the coding sequence of the extended cDNA or portion thereof is inserted in frame with the gene encoding the other half of

-51-

the chimera. The other half of the chimera may be β-globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to β-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites may be engineered between the β-globin gene or the nickel binding polypeptide and the extended cDNA or portion thereof. Thus, the two polypeptides of the chimera may be separated from one another by
5   protease digestion.

One useful expression vector for generating β-globin chimerics is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al.,
10   (**Basic Methods in Molecular Biology**, L.G. Davis, M.D. Dibner, and J.F. Battey, ed., Elsevier Press, NY, 1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using in vitro translation systems such as the In vitro Express™ Translation Kit (Stratagene).

Following expression and purification of the secreted proteins encoded by the 5' ESTs, extended cDNAs, or
15   fragments thereof, the purified proteins may be tested for the ability to bind to the surface of various cell types as described in Example 31 below. It will be appreciated that a plurality of proteins expressed from these cDNAs may be included in a panel of proteins to be simultaneously evaluated for the activities specifically described below, as well as other biological roles for which assays for determining activity are available.

EXAMPLE 31

20   ### Analysis of Secreted Proteins to Determine Whether they Bind to the Cell Surface

The proteins encoded by the 5' ESTs, extended cDNAs, or fragments thereof are cloned into expression vectors such as those described in Example 30. The proteins are purified by size, charge, immunochromatography or other techniques familiar to those skilled in the art. Following purification, the proteins are labeled using techniques known to those skilled in the art. The labeled proteins are incubated with cells or cell lines derived from a variety of organs or
25   tissues to allow the proteins to bind to any receptor present on the cell surface. Following the incubation, the cells are washed to remove non-specifically bound protein. The labeled proteins are detected by autoradiography. Alternatively, unlabeled proteins may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto.

Specificity of cell surface binding may be analyzed by conducting a competition analysis in which various
30   amounts of unlabeled protein are incubated along with the labeled protein. The amount of labeled protein bound to the cell surface decreases as the amount of competitive unlabeled protein increases. As a control, various amounts of an unlabeled protein unrelated to the labeled protein is included in some binding reactions. The amount of labeled protein bound to the cell surface does not decrease in binding reactions containing increasing amounts of unrelated unlabeled protein, indicating that the protein encoded by the cDNA binds specifically to the cell surface.

·52·

As discussed above, secreted proteins have been shown to have a number of important physiological effects and, consequently, represent a valuable therapeutic resource. The secreted proteins encoded by the extended cDNAs or portions thereof made according to Examples 27-29 may be evaluated to determine their physiological activities as described below.

5 **EXAMPLE 32**

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Cytokine, Cell Proliferation or Cell Differentiation Activity

As discussed above, secreted proteins may act as cytokines or may affect cellular proliferation or differentiation. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or
10 more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B5, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c and CMK. The proteins encoded by the above extended cDNAs or portions thereof may be evaluated for their ability to regulate T cell or thymocyte proliferation
15 in assays such as those described above or in the following references: Current Protocols in Immunology, Ed. by J.E. Coligan et al., Greene Publishing Associates and Wiley-Interscience; Takai et al. J. Immunol. 137:3494-3500, 1986. Bertagnolli et al. J. Immunol. 145:1706-1712, 1990. Bertagnolli et al., Cellular Immunology 133:327-341, 1991. Bertagnolli, et al. J. Immunol. 149:3778-3783, 1992; Bowman et al., J. Immunol. 152:1756-1761, 1994.

In addition, numerous assays for cytokine production and/or the proliferation of spleen cells, lymph node cells
20 and thymocytes are known. These include the techniques disclosed in Current Protocols in Immunology. J.E. Coligan et al. Eds., Vol 1 pp. 3.12.1-3.12.14 John Wiley and Sons, Toronto. 1994; and Schreiber, R.D. Current Protocols in Immunology., supra Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

The proteins encoded by the cDNAs may also be assayed for the ability to regulate the proliferation and differentiation of hematopoietic or lymphopoietic cells. Many assays for such activity are familiar to those skilled in the
25 art, including the assays in the following references: Bottomly, K., Davis, L.S. and Lipsky, P.E., Measurement of Human and Murine Interleukin 2 and Interleukin 4, Current Protocols in Immunology., J.E. Coligan et al. Eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 36:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Nordan, R., Measurement of Mouse and Human Interleukin 6 Current Protocols in Immunology. J.E. Coligan et al. Eds. Vol 1 pp. 6.6.1-6.6.5,
30 John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Bennett, F., Giannotti, J., Clark, S.C. and Turner, K.J., Measurement of Human Interleukin 11 Current Protocols in Immunology. J.E. Coligan et al. Eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto 1991; Ciarletta. A., Giannotti, J., Clark, S.C. and Turner, K.J., Measurement of Mouse and Human Interleukin 9 Current Protocols in Immunology. J.E. Coligan et al., Eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

-53-

The proteins encoded by the cDNAs may also be assayed for their ability to regulate T-cell responses to

antigens. Many assays for such activity are familiar to those skilled in the art, including the assays described in the

following references: Chapter 3 (In Vitro Assays for Mouse Lymphocyte Function), Chapter 6 (Cytokines and Their

Cellular Receptors) and Chapter 7, (Immunologic Studies in Humans) in **Current Protocols in Immunology**, J.E. Coligan

5 et al. Eds. Greene Publishing Associates and Wiley-Interscience; Weinberger et al., **Proc. Natl. Acad. Sci. USA**

77:6091-6095, 1980; Weinberger et al., **Eur. J. Immun.** 11:405-411, 1981; Takai et al., **J. Immunol.** 137:3494-3500,

1986; Takai et al., **J. Immunol.** 140:508-512, 1988.

Those proteins which exhibit cytokine, cell proliferation, or cell differentiation activity may then be formulated

as pharmaceuticals and used to treat clinical conditions in which induction of cell proliferation or differentiation is

10 beneficial. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the

expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the

proteins as desired.

EXAMPLE 33

Assaying the Proteins Expressed from Extended cDNAs or Portions

15   Thereof for Activity as Immune System Regulators

The proteins encoded by the cDNAs may also be evaluated for their effects as immune regulators. For

example, the proteins may be evaluated for their activity to influence thymocyte or splenocyte cytotoxicity. Numerous

assays for such activity are familiar to those skilled in the art including the assays described in the following references:

Chapter 3 (In Vitro Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic studies in Humans) in

20 **Current Protocols in Immunology**, J.E. Coligan et al. Eds. Greene Publishing Associates and Wiley-Interscience;

Herrmann et al., **Proc. Natl. Acad. Sci. USA** 78:2488-2492, 1981; Herrmann et al., **J. Immunol.** 128:1968-1974,

1982; Handa et al., **J. Immunol.** 135:1564-1572, 1985; Takai et al., **J. Immunol.** 137:3494-3500, 1986; Takai et al.,

J. Immunol. 140:508-512, 1988; Herrmann et al., **Proc. Natl. Acad. Sci. USA** 78:2488-2492, 1981; Herrmann et al.,

J. Immunol. 128:1968-1974, 1982; Handa et al., **J. Immunol.** 135:1564-1572, 1985; Takai et al., **J. Immunol.**

25 137:3494-3500, 1986; Bowman et al., **J. Virology** 61:1992-1998; Takai et al., **J. Immunol.** 140:508-512, 1988;

Bertagnolli et al., **Cellular Immunology** 133:327-341, 1991; Brown et al., **J. Immunol.** 153:3079-3092, 1994.

The proteins encoded by the cDNAs may also be evaluated for their effects on T-cell dependent immunoglobulin

responses and isotype switching. Numerous assays for such activity are familiar to those skilled in the art, including the

assays disclosed in the following references: Maliszewski, **J. Immunol.** 144:3028-3033, 1990; Mond, J.J. and

30 Brunswick, M Assays for B Cell Function: *In vitro* Antibody Production, Vol 1 pp. 3.8.1-3.8.16 in **Current Protocols in

Immunology**. J.E. Coligan et al Eds., John Wiley and Sons, Toronto. 1994.

The proteins encoded by the cDNAs may also be evaluated for their effect on immune effector cells, including

their effect on Th1 cells and cytotoxic lymphocytes. Numerous assays for such activity are familiar to those skilled in

the art, including the assays disclosed in the following references: Chapter 3 (In Vitro Assays for Mouse Lymphocyte

-54-

Function 3.1-3.19) and Chapter 7 (Immunologic Studies in Humans) in Current Protocols in Immunology, J.E. Coligan et al. Eds., Greene Publishing Associates and Wiley-Interscience; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al.; J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

The proteins encoded by the cDNAs may also be evaluated for their effect on dendritic cell mediated activation
5   of naive T-cells. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264,
10   1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

The proteins encoded by the cDNAs may also be evaluated for their influence on the lifetime of lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca
15   et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular immunology 155:111-122,
20   1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

Those proteins which exhibit activity as immune system regulators activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of immune activity is beneficial. For example, the protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as
25   effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present
30   invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis,

-55-

myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present
invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic
asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for
example, organ transplantation), may also be treatable using a protein of the present invention.

5   Using the proteins of the invention it may also be possible to regulate immune responses, in a number of ways.
Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve
preventing the induction of an immune response. The functions of activated T-cells may be inhibited by suppressing T
cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an
active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent.

10   Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in
that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally,
tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the
tolerizing agent.

          Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte
15   antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will
be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example,
blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue
transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune
reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7
20   lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-
2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen
(e.g., B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural
ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte
antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an
25   immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing
tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the
necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance
in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

          The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed
30   using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used
include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used
to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science
257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA. 89:11102-11105 (1992). In addition, murine models

-56-

of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function in vivo on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which
5 promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead
10 to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/pr/pr mice or NZB hybrid mice, murine autoimmuno collagen arthritis, diabetes mellitus in OD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

15 Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory
20 form of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to T cells
25 in vivo, thereby activating the T cells.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be
30 transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected ex vivo with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection in vivo.

·57·

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic

5 acids encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I α chain protein and β_2 macroglobulin protein or an MHC class II α chain protein and an MHC class II β chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class II or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression

10 of an MHC class II associated protein, such as the invariant chain,can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells

15 to increase or decrease the expression of the proteins as desired.

EXAMPLE 34

Assaying the Proteins Expressed from Extended cDNAs
or Portions Thereof for Hematopoiesis Regulating Activity

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their hematopoiesis

20 regulating activity. For example, the effect of the proteins on embryonic stem cell differentiation may be evaluated. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Johansson et al. **Cellular Biology** 15:141-151, 1995; Keller et al., **Molecular and Cellular Biology** 13:473-486, 1993; McClanahan et al., **Blood** 81:2903-2915, 1993.

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their influence on

25 the lifetime of stem cells and stem cell differentiation. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Freshney, M.G. Methylcellulose Colony Forming Assays, in **Culture of Hematopoietic Cells**. R.I. Freshney, et al. Eds. pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama et al., **Proc. Natl. Acad. Sci. USA** 89:5907-5911, 1992; McNiece, I.K. and Briddell, R.A. Primitive Hematopoietic Colony Forming Cells with High Proliferative Potential, in **Culture of Hematopoietic Cells**. R.I.

30 Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, NY. 1994; Neben et al., **Experimental Hematology** 22:353-359, 1994; Ploemacher, R.E. Cobblestone Area Forming Cell Assay, In **Culture of Hematopoietic Cells**. R.I. Freshney, et al. Eds. pp. 1-21, Wiley-Liss, Inc., New York, NY. 1994; Spooncer, E., Dexter, M. and Allen, T. Long Term Bone Marrow Cultures in the Presence of Stromal Cells, in **Culture of Hematopoietic Cells**. R.I. Freshney, et al. Eds.

-58-

pp. 163-179, Wiley-Liss, Inc., New York, NY. 1994; and Sutherland, H.J. Long Term Culture Initiating Cell Assay, in Culture of Hematopoietic Cells. R.I. Freshney, et al. Eds. pp. 139-162, Wiley-Liss, Inc., New York, NY. 1994.

Those proteins which exhibit hematopoiesis regulatory activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of hematopoeisis is beneficial. For example, a protein of the present
5   invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid
10 cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem
15 cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantion, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or
20 genetically manipulated for gene therapy. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 35

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Tissue Growth

25   The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their effect on tissue growth. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in International Patent Publication No. WO95/16035, International Patent Publication No. WO95/05846 and International Patent Publication No. WO91/07491.

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound
30 Healing, pps. 71-112 (Maibach, H1 and Rovee, DT, eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

Those proteins which are involved in the regulation of tissue growth may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of tissue growth is beneficial. For example, a protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or

-59-

nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and

5 other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair

10 processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

15 Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to

20 tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate

25 growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

30 The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e., for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as

-60-

Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

5   Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium) muscle

10   (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to generate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokinc damage.

15   A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

20   **EXAMPLE 36**

Assaying the Proteins Expressed from Extended cDNAs or Portions
Thereof for Regulation of Reproductive Hormones or Cell Movement

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their ability to regulate reproductive hormones, such as follicle stimulating hormone. Numerous assays for such activity are familiar to

25   those skilled in the art, including the assays disclosed in the following references: Vale et al., **Endocrinology** 91:562-572, 1972; Ling et al., **Nature** 321:779-782, 1986; Vale et al., **Nature** 321:776-779, 1986; Mason et al., **Nature** 318:659-663, 1985; Forage et al., **Proc. Natl. Acad. Sci. USA** 83:3091-3095, 1986. Chapter 6.12 (Measurement of Alpha and Beta Chemokines) **Current Protocols in Immunology,** J.E. Coligan et al. Eds. Greene Publishing Associates and Wiley-Intersciece; Taub et al. **J. Clin. Invest.** 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller

30   et al. **Eur. J. Immonol.** 25:1744-1748; Gruber et al. **J. of Immunol.** 152:5860-5867, 1994; Johnston et al. **J. of Immunol.** 153:1762-1768, 1994.

Those proteins which exhibit activity as reproductive hormones or regulators of cell movement may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of reproductive hormones or cell movement are beneficial. For example, a protein of the present invention may also exhibit activin- or inhibin-related

-61-

activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while

activins are characterized by their ability to stimulate the release of folic stimulating hormone (FSH). Thus, a protein of

the present invention, alone or in heterodimers with a member of the inhibin α family, may be useful as a contraceptive

based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals.

5   Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein

of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-B group, may be useful as

a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the

anterior pituitary. See, for example, United States Patent 4,798,885. A protein of the invention may also be useful for

advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive

10 performance of domestic animals such as cows, sheep and pigs.

Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the

expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the

proteins as desired.

EXAMPLE 36A

15   Assaying the Proteins Expressed from Extended cDNAs or

Portions Thereof for Chemotactic/Chemokinetic Activity

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for chemotactic/chemokinetic

activity. For example, a protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a

chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells,

20 eosinophils, epithelial and/or endothelial cells. Chemotactic and chmokinetic proteins can be used to mobilize or attract a

desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in

treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example,

attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune

responses against the tumor or infecting agent.

25   A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or

indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability

to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population

of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

30 Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that

measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to

induce the adhension of one cell population to another cell population. Suitable assays for movement and adhesion

include, without limitation, those described in: Current Protocols in Immunology, Ed by J.E. Coligan, A.M. Kruisbeek,

D.H. Margulies, E.M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12,

Measurement of alpha and beta Chemokincs 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Mueller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol, 153:1762-1768, 1994.

EXAMPLE 37

5   Assaying the Proteins Expressed from Extended cDNAs or

Portions Thereof for Regulation of Blood Clotting

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their effects on blood clotting. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references: Linet et al., **J. Clin. Pharmacol.** 26:131-140, 1986; Burdick et al., **Thrombosis Res.**

10   45:413-419, 1987; Humphrey et al., **Fibrinolysis** 5:71-79 (1991); Schaub, **Prostaglandins** 35:467-474, 1988.

Those proteins which are involved in the regulation of blood clotting may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of blood clotting is beneficial. For example, a protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulations disorders (including hereditary disorders, such as hemophilias) or to

15 enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as,for example, infarction of cardiac and central nervous system vessels (e.g., stroke). Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the

20 expression of the proteins as desired.

EXAMPLE 38

Assaying the Proteins Expressed from Extended cDNAs or

Portions Thereof for Involvement in Receptor/Ligand Interactions

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for their involvement

25 in receptor/ligand interactions. Numerous assays for such involvement are familiar to those skilled in the art, including the assays disclosed in the following references: Chapter 7.28 (Measurement of Cellular Adhesion under Static Conditions 7.28.1-7.28.22) in **Current Protocols in Immunology**, J.E. Coligan et al. Eds. Greene Publishing Associates and Wiley-Interscience; Takai et al., **Proc. Natl. Acad. Sci.** USA 84:6864-6868, 1987; Bierer et al., **J. Exp. Med.** 168:1145-1156, 1988; Rosenstein et al., **J. Exp. Med.** 169:149-160, 1989; Stoltenborg et al., **J. Immunol. Methods**

30 175:59-68, 1994; Stitt et al., **Cell** 80:661-670, 1995; Gyuris et al., *Cell* 75:791-803, 1993.

For example, the proteins of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion

-63-

molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune respones). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as
5 inhibitors of receptor/ligand interactions.

EXAMPLE 38A

Assaying the Proteins Expressed from Extended cDNAs or Portions
Thereof for Anti-Inflammatory Activity

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for anti-inflammatory
10 activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including
15 without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)),ischemia-reperfusioninury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

20 -   **EXAMPLE 38B**

Assaying the Proteins Expressed from Extended cDNAs or
Portions Thereof for Tumor Inhibition Activity

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for tumor inhibition activity. In addition to the activities described above for immunological treatment or prevention of tumors, a protein of
25 the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

30 A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or

-64-

circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent

5   behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such

10   protein or another material or entity which is cross-reactive with such protein.

<div align="center">

EXAMPLE 39

Identification of Proteins which Interact with

Polypeptides Encoded by Extended cDNAs

</div>

Proteins which interact with the polypeptides encoded by extended cDNAs or portions thereof, such as

15   receptor proteins, may be identified using two hybrid systems such as the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech), the extended cDNAs or portions thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. cDNAs in a cDNA library which encode proteins which might interact with the polypeptides encoded by the extended cDNAs or portions thereof

20   are inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain plasmids encoding

25   proteins which interact with the polypeptide encoded by the extended cDNAs or portions thereof.

Alternatively, the system described in Lustig et al., Methods in Enzymology 283: 83-99 (1997) may be used for identifying molecules which interact with the polypeptides encoded by extended cDNAs. In such systems, *in vitro* transcription reactions are performed on a pool of vectors containing extended cDNA inserts cloned downstream of a promoter which drives *in vitro* transcription. The resulting pools of mRNAs are introduced into *Xenopus laevis* oocytes.

30   The oocytes are then assayed for a desired acitivity.

Alternatively, the pooled *in vitro* transcription products produced as described above may be translated *in vitro*. The pooled *in vitro* translation products can be assayed for a desired activity or for interaction with a known polypeptide.

Proteins or other molecules interacting with polypeptides encoded by extended cDNAs can be found by a variety of additional techniques. In one method, affinity columns containing the polypeptide encoded by the extended cDNA or a portion thereof can be constructed. In some versions, of this method the affinity column contains chimeric proteins in which the protein encoded by the extended cDNA or a portion thereof is fused to glutathione S-transferase.

5   A mixture of cellular proteins or pool of expressed proteins as described above and is applied to the affinity column. Proteins interacting with the polypeptide attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen et al. Electrophoresis, 18, 588-598 (1997). Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

10   Proteins interacting with polypeptides encoded by extended cDNAs or portions thereof can also be screened by using an Optical Biosensor as described in Edwards & Leatherbarrow, Analytical Biochemistry, 246, 1-6 (1997). The main advantage of the method is that it allows the determination of the association rate between the protein and other interacting molecules. Thus, it is possible to specifically select interacting molecules with a high or low association rate. Typically a target molecule is linked to the sensor surface (through a carboxymethl dextran matrix) and a sample of test

15   molecules is placed in contact with the target molecules. The binding of a test molecule to the target molecule causes a change in the refractive index and/ or thickness. This change is detected by the Biosensor provided it occurs in the evanescent field (which extend a few hundred manometers from the sensor surface). In these screening assays, the target molecule can be one of the polypeptides encoded by extended cDNAs or a portion thereof and the test sample can be a collection of proteins extracted from tissues or cells, a pool of expressed proteins, combinatorial peptide and/ or

20 .   chemical libraries,or phage displayed peptides. The tissues or cells from which the test proteins are extracted can originate from any species.

In other methods, a target protein is immobilized and the test population is a collection of unique polypeptides encoded by the extended cDNAs or portions thereof.

To study the interaction of the proteins encoded by the extended cDNAs or portions thereof with drugs, the

25   microdialysis coupled to HPLC method described by Wang et al., Chromatographia, 44, 205-208(1997) or the affinity capillary electrophoresis method described by Busch et al., J. Chromatogr. 777:311-328 (1997), the disclosures of which are incorporated herein by referenc can be used.

The system described in U.S. Patent No. 5,654,150 may also be used to identify molecules which interact with the polypeptides encoded by the extended cDNAs. In this system, pools of extended cDNAs are transcribed and

30   translated in vitro and the reaction products are assayed for interaction with a known polypeptide or antibody.

It will be appreciated by those skilled in the art that the proteins expressed from the extended cDNAs or portions may be assayed for numerous activities in addition to those specifically enumerated above. For example, the expressed proteins may be evaluated for applications involving control and regulation of inflammation, tumor

-66-

proliferation or metastasis, infection, or other clinical conditions. In addition, the proteins expressed from the extended cDNAs or portions thereof may be useful as nutritional agents or cosmetic agents.

The proteins expressed from the extended cDNAs or portions thereof may be used to generate antibodies capable of specifically binding to the expressed protein or fragments thereof as described in Example 40 below. The

5   antibodies may capable of binding a full length protein encoded by one of the sequences of SEQ ID NOs. 40-140 and 242-377, a mature protein encoded by one of the sequences of SEQ ID NOs. 40-140 and 242-377, or a signal peptide encoded by one of the sequences of SEQ ID Nos. 40-140 and 242-377. Alternatively, the antibodies may be capable of binding fragments of the proteins expressed from the extended cDNAs which comprise at least 10 amino acids of the sequences of SEQ ID NOs: 141-241 and 378-513. In some embodiments, the antibodies may be capable of binding

10   fragments of the proteins expressed from the extended cDNAs which comprise at least 15 amino acids of the sequences of SEQ ID NOs: 141-241 and 378-513. In other embodiments, the antibodies may be capable of binding fragments of the proteins expressed from the extended cDNAs which comprise at least 25 amino acids of the sequences of SEQ ID NOs: 141-241 and 378-513. In further embodiments, the antibodies may be capable of binding fragments of the proteins expressed from the extended cDNAs which comprise at least 40 amino acids of the sequences of SEQ ID NOs:

15   141-241 and 378-513.

EXAMPLE 40

Production of an Antibody to a Human Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells as described in Example 30. The concentration of protein in the final preparation is adjusted, for example, by concentration on an

20   Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A.   Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., Nature 256:495 (1975) or

25   derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein or peptides derived therefrom over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the

30   culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as Elisa, as originally described by Engvall, E., Meth. Enzymol. 70:419 (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. **Basic Methods in Molecular Biology** Elsevier, New York. Section 21-2.

-67-

B.   **Polyclonal Antibody Production by Immunization**

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein or peptides derived therefrom described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors

5   related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. **J. Clin. Endocrinol. Metab. 33:988-991 (1971).**

10   Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: **Handbook of Experimental Immunology** D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as

15   described, for example, by Fisher, D., Chap. 42 in: **Manual of Clinical Immunology,** 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic

20   compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

V. Use of Extended cDNAs or Portions Thereof as Reagents

The extended cDNAs of the present invention may be used as reagents in isolation procedures, diagnostic assays, and forensic procedures. For example, sequences from the extended cDNAs (or genomic DNAs obtainable

25   therefrom) may be detectably labeled and used as probes to isolate other sequences capable of hybridizing to them. In addition, sequences from the extended cDNAs (or genomic DNAs obtainable therefrom) may be used to design PCR primers to be used in isolation, diagnostic, or forensic procedures.

EXAMPLE 41

Preparation of PCR Primers and Amplification of DNA

30   The extended cDNAs (or genomic DNAs obtainable therefrom) may be used to prepare PCR primers for a variety of applications, including isolation procedures for cloning nucleic acids capable of hybridizing to such sequences, diagnostic techniques and forensic techniques. The PCR primers are at least 10 bases, and preferably at least 12, 15, or 17 bases in length. More preferably, the PCR primers are at least 20-30 bases in length. In some embodiments, the PCR primers may be more than 30 bases in length. It is preferred that the primer pairs have approximately the same G/C

-68-

ratio, so that melting temperatures are approximately the same. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering White, B.A. Ed. in Methods in Molecular Biology 67: Humana Press, Totowa 1997. In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with

5   dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites.

10   EXAMPLE 42

Use of Extended cDNAs as Probes

Probes derived from extended cDNAs or portions thereof (or genomic DNAs obtainable therefrom) may be labeled with detectable labels familiar to those skilled in the art, including radioisotopes and non-radioactive labels, to provide a detectable probe. The detectable probe may be single stranded or double stranded and may be made using

15   techniques known in the art, including in vitro transcription, nick translation, or kinase reactions. A nucleic acid sample containing a sequence capable of hybridizing to the labeled probe is contacted with the labeled probe. If the nucleic acid in the sample is double stranded, it may be denatured prior to contacting the probe. In some applications, the nucleic acid sample may be immobilized on a surface such as a nitrocellulose or nylon membrane. The nucleic acid sample may comprise nucleic acids obtained from a variety of sources, including genomic DNA, cDNA libraries, RNA, or tissue

20   samples.

Procedures used to detect the presence of nucleic acids capable of hybridizing to the detectable probe include well known techniques such as Southern blotting, Northern blotting, dot blotting, colony hybridization, and plaque hybridization. In some applications, the nucleic acid capable of hybridizing to the labeled probe may be cloned into vectors such as expression vectors, sequencing vectors, or in vitro transcription vectors to facilitate the characterization

25   and expression of the hybridizing nucleic acids in the sample. For example, such techniques may be used to isolate and clone sequences in a genomic library or cDNA library which are capable of hybridizing to the detectable probe as described in Example 30 above.

PCR primers made as described in Example 41 above may be used in forensic analyses, such as the DNA fingerprinting techniques described in Examples 43-47 below. Such analyses may utilize detectable probes or primers

30   based on the sequences of the extended cDNAs isolated using the 5' ESTs (or genomic DNAs obtainable therefrom).

EXAMPLE 43

Forensic Matching by DNA Sequencing

In one exemplary method, DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers based on a number of the extended cDNAs (or

-69-

genomic DNAs obtainable therefrom), is then utilized in accordance with Example 41 to amplify DNA of approximately
100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a test subject. Each
of these identification DNAs is then sequenced using standard techniques, and a simple database comparison determines
the differences, if any, between the sequences from the subject and those from the sample. Statistically significant

5   differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This
lack of identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated
with a large number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences of 100
bases in length are used to prove identity between the suspect and the sample.

EXAMPLE 44

10   Positive Identification by DNA Sequencing

The technique outlined in the previous example may also be used on a larger scale to provide a unique
fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of sequences
from Table IV and the appended sequence listing. Preferably, 20 to 50 different primers are used. These primers are
used to obtain a corresponding number of PCR-generated DNA segments from the individual in question in accordance

15   with Example 41. Each of these DNA segments is sequenced, using the methods set forth in Example 43. The database
of sequences generated through this procedure uniquely identifies the individual from whom the sequences were
obtained. The same panel of primers may then be used at any later time to absolutely correlate tissue or other biological
specimen with that individual.

EXAMPLE 45

20   Southern Blot Forensic Identification

The procedure of Example 44 is repeated to obtain a panel of at least 10 amplified sequences from an
individual and a specimen. Preferably, the panel contains at least 50 amplified sequences. More preferably, the panel
contains 100 amplified sequences. In some embodiments, the panel contains 200 amplified sequences. This PCR-
generated DNA is then digested with one or a combination of, preferably, four base specific restriction enzymes. Such

25   enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments
are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting
techniques well known to those with skill in the art. For a review of Southern blotting see Davis et al. (**Basic Methods
in Molecular Biology**, 1986, Elsevier Press. pp 62-65).

A panel of probes based on the sequences of the extended cDNAs (or genomic DNAs obtainable therefrom), or

30   fragments thereof of at least 10 bases, are radioactively or colorimetrically labeled using methods known in the art, such
as nick translation or end labeling, and hybridized to the Southern blot using techniques known in the art (Davis et al.,
supra). Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the extended cDNA (or
genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from
the extended cDNA (or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30

-70-

nucleotides from the extended cDNA (or genomic DNAs obtainable therefrom). In other embodiments, the probe comprises at least 40, at least 50, at least 75, at least 100, at least 150, or at least 200 consecutive nucleotides from the extended cDNA (or genomic DNAs obtainable therefrom).

Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of extended cDNAs (or genomic DNAs obtainable therefrom) will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of extended cDNA probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

EXAMPLE 46

Dot Blot Identification Procedure

Another technique for identifying individuals using the extended cDNA sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length are synthesized that correspond to at least 10, preferably 50 sequences from the extended cDNAs or genomic DNAs obtainable therefrom. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labeled with P^{32} using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting the genomic DNA onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the genomic sequences is baked or UV linked to the filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis et al. supra). The ^{32}P labeled DNA fragments are sequentially hybridized with successively stringent conditions to detect minimal differences between the 30 bp sequence and the DNA. Tetramethylammonium chloride is useful for identifying clones containing small numbers of nucleotide mismatches (Wood et al., **Proc. Natl. Acad. Sci. USA 82(6):1585-1588 (1985)**). A unique pattern of dots distinguishes one individual from another individual.

Extended cDNAs or oligonucleotides containing at least 10 consecutive bases from these sequences can be used as probes in the following alternative fingerprinting technique. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the extended cDNA (or genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from the extended cDNA (or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30 nucleotides from the extended cDNA (or genomic DNAs obtainable therefrom). In other embodiments, the probe comprises at least 40, at least 50, at least 75, at least 100, at least 150, or at least 200 consecutive nucleotides from the extended cDNA (or genomic DNAs obtainable therefrom).

-71-

Preferably, a plurality of probes having sequences from different genes are used in the alternative fingerprinting technique. Example 47 below provides a representative alternative fingerprinting procedure in which the probes are derived from extended cDNAs.

EXAMPLE 47

5

Alternative "Fingerprint" Identification Technique

20-mer oligonucleotides are prepared from a large number, e.g. 50, 100, or 200, of extended cDNA sequences (or genomic DNAs obtainable therefrom) using commercially available oligonucleotide services such as Genset, Paris, France. Cell samples from the test subject are processed for DNA using techniques well known to those with skill in the art. The nucleic acid is digested with restriction enzymes such as EcoRI and XbaI. Following digestion, samples are

10 applied to wells for electrophoresis. The procedure, as known in the art, may be modified to accommodate polyacrylamide electrophoresis, however in this example, samples containing 5 ug of DNA are loaded into wells and separated on 0.8% agarose gels. The gels are transferred onto nitrocellulose using standard Southern blotting techniques.

10 ng of each of the oligonucleotides are pooled and end-labeled with P^{32}. The nitrocellulose is prehybridized

15 with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

It is additionally contemplated within this example that the number of probe sequences used can be varied for additional accuracy or clarity.

The antibodies generated in Examples 30 and 40 above may be used to identify the tissue type or cell species

20 from which a sample is derived as described above.

EXAMPLE 48

Identification of Tissue Types or Cell Species by Means of

Labeled Tissue Specific Antibodies

Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of

25 antibody preparations according to Examples 30 and 40 which are conjugated, directly or indirectly, to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

Antisera for these procedures must have a potency exceeding that of the native preparation, and for that

30 reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

-72-

A.   Immunohistochemical Techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, H., Chap. 26 in: **Basic 503 Clinical Immunology**, 3rd Ed. Lange, Los Altos, California (1980) or Rose, N. et al., Chap. 12 in: **Methods in Immunodiagnosis**, 2d Ed. John Wiley 503 Sons, New

5   York (1980).

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific antitissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody

10   complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ^{125}I, and detected by overlaying the antibody treated preparation with photographic emulsion.

Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single protein or peptide identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required.

15   Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 μm, unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer.

20   Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed.

If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such

25   labeled sera are commercially available.

The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

B.   Identification of Tissue Specific Soluble Proteins

The visualization of tissue specific proteins and identification of unknown tissues from that procedure is

30   carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection.

A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice

-73-

in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction concentrated if necessary and reserved for analysis.

A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide electrophoresis as described, for example, by Davis, L. et al., Section 19-2 in: **Basic Methods in**
5   **Molecular Biology** (P. Leder, ed), Elsevier, New York (1986), using a range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5 to 55 μl, and containing from about 1 to 100 μg protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies
10   are prepared. The procedure, known as Western Blot Analysis, is well described in Davis, L. et al., (above) Section 19-3.

One set of nitrocellulose blots is stained with Coomassie Blue dye to visualize the entire set of proteins for comparison with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described in Examples 30 and 40. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.
15   In either procedure A or B, a detectable label can be attached to the primary tissue antigen-primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin conjugated marker. According to yet another strategy, enzyme labeled or radioactive
20   protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody.

The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from extended cDNA sequences, can identify tissues of unknown origin, for example, forensic samples, or differentiated tumor tissue that has metastasized to foreign
25   bodily sites.

In addition to their applications in forensics and identification, extended cDNAs (or genomic DNAs obtainable therefrom) may be mapped to their chromosomal locations. Example 49 below describes radiation hybrid (RH) mapping of human chromosomal regions using extended cDNAs. Example 50 below describes a representative procedure for mapping an extended cDNA (or a genomic DNA obtainable therefrom) to its location on a human chromosome. Example
30   51 below describes mapping of extended cDNAs (or genomic DNAs obtainable therefrom) on metaphase chromosomes by Fluorescence In Situ Hybridization (FISH).

EXAMPLE 49

Radiation hybrid mapping of Extended cDNAs to the human genome

-74-

Radiation hybrid (RH) mapping is a somatic cell genetic approach that can be used for high resolution mapping of the human genome. In this approach, cell lines containing one or more human chromosomes are lethally irradiated, breaking each chromosome into fragments whose size depends on the radiation dose. These fragments are rescued by fusion with cultured rodent cells, yielding subclones containing different portions of the human genome. This technique

5   is described by Benham et al. (*Genomics* 4:509-517, 1989) and Cox et al., (*Science* 250:245-250, 1990). The random and independent nature of the subclones permits efficient mapping of any human genome marker. Human DNA isolated from a panel of 80-100 cell lines provides a mapping reagent for ordering extended cDNAs (or genomic DNAs obtainable therefrom). In this approach, the frequency of breakage between markers is used to measure distance, allowing construction of fine resolution maps as has been done using conventional ESTs (Schuler et al., *Science* 274:540-546,

10   1996).

RH mapping has been used to generate a high-resolution whole genome radiation hybrid map of human chromosome 17q22-q25.3 across the genes for growth hormone (GH) and thyridine kinase (TK) (Foster et al., *Genomics* 33:185-192, 1996), the region surrounding the Gorlin syndrome gene (Obermayr et al., *Eur. J. Hum. Genet.* 4:242-245, 1996), 60 loci covering the entire short arm of chromosome 12 (Raeymaekers et al., *Genomics* 29:170-178, 1995), the

15   region of human chromosome 22 containing the neurofibromatosis type 2 locus (Frazer et al., *Genomics* 14:574-584, 1992) and 13 loci on the long arm of chromosome 5 (Warrington et al., *Genomics* 11:701-708, 1991).

EXAMPLE 50

Mapping of Extended cDNAs to Human

Chromosomes using PCR techniques

20   Extended cDNAs (or genomic DNAs obtainable therefrom) may be assigned to human chromosomes using PCR based methodologies. In such approaches, oligonucleotide primer pairs are designed from the extended cDNA sequence (or the sequence of a genomic DNA obtainable therefrom) to minimize the chance of amplifying through an intron. Preferably, the oligonucleotide primers are 18-23 bp in length and are designed for PCR amplification. The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see

25   Erlich, H.A., **PCR Technology: Principles and Applications for DNA Amplification**. 1992. W.H. Freeman and Co., New York.

The primers are used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions are as follows: 60 ng of genomic DNA is used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1 μCu of a ^{32}P-labeled deoxycytidine triphosphate. The PCR is

30   performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94°C, 1.4 min; 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products are analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the length of the resulting PCR product is identical to the distance between the ends of the primer sequences in the extended cDNA from which the primers are derived, then the PCR reaction is repeated with DNA templates from two panels of human-rodent somatic cell hybrids, BIOS

-75-

PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

PCR is used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given extended cDNA (or genomic DNA obtainable therefrom). DNA is isolated from the somatic

5 hybrids and used as starting templates for PCR reactions using the primer pairs from the extended cDNAs (or genomic DNAs obtainable therefrom). Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the extended cDNA (or genomic DNA obtainable therefrom) will yield an amplified fragment. The extended cDNAs (or genomic DNAs obtainable therefrom) are assigned to a chromosome by analysis of the segregation pattern of PCR products from the somatic hybrid DNA templates. The single human chromosome present in all cell

10 hybrids that give rise to an amplified fragment is the chromosome containing that extended cDNA (or genomic DNA obtainable therefrom). For a review of techniques and analysis of results from somatic cell gene mapping experiments. (See Ledbetter et al., **Genomics** 6:475-481 (1990).)

Alternatively, the extended cDNAs (or genomic DNAs obtainable therefrom) may be mapped to individual chromosomes using FISH as described in Example 51 below.

15   **EXAMPLE 51**

                                     Mapping of Extended 5' ESTs to Chromosomes

                                     Using Fluorescence in situ Hybridization

Fluorescence in situ hybridization allows the extended cDNA (or genomic DNA obtainable therefrom) to be mapped to a particular location on a given chromosome. The chromosomes to be used for fluorescence in situ

20 hybridization techniques may be obtained from a variety of sources including cell cultures, tissues, or whole blood.

In a preferred embodiment, chromosomal localization of an extended cDNA (or genomic DNA obtainable therefrom) is obtained by FISH as described by Cherif et al. (*Proc. Natl. Acad. Sci. U.S.A.*, 87:6639-6643, 1990). Metaphase chromosomes are prepared from phytohemagglutinin (PHA)-stimulated blood cell donors. PHA-stimulated lymphocytes from healthy males are cultured for 72 h in RPMI-1640 medium. For synchronization, methotrexate (10

25 μM) is added for 17 h, followed by addition of 5-bromodeoxyuridine (5-BudR, 0.1 mM) for 6 h. Colcemid (1 μg/ml) is added for the last 15 min before harvesting the cells. Cells are collected, washed in RPMI, incubated with a hypotonic solution of KCl (75 mM) at 37°C for 15 min and fixed in three changes of methanol:acetic acid (3:1). The cell suspension is dropped onto a glass slide and air dried. The extended cDNA (or genomic DNA obtainable therefrom) is labeled with biotin-16 dUTP by nick translation according to the manufacturer's instructions (Bethesda Research

30 Laboratories, Bethesda, MD), purified using a Sephadex G-50 column (Pharmacia, Upssala, Sweden) and precipitated. Just prior to hybridization, the DNA pellet is dissolved in hybridization buffer (50% formamide, 2 X SSC, 10% dextran sulfate, 1 mg/ml sonicated salmon sperm DNA, pH 7) and the probe is denatured at 70°C for 5-10 min.

Slides kept at -20°C are treated for 1 h at 37°C with RNase A (100 μg/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol series. Chromosome preparations are denatured in 70% formamide, 2 X SSC for 2 min at

-76-

70°C, then dehydrated at 4°C. The slides are treated with proteinase K (10 μg/100 ml in 20 mM Tris-HCl, 2 mM CaCl₂)

at 37°C for 8 min and dehydrated. The hybridization mixture containing the probe is placed on the slide, covered with a

coverslip, sealed with rubber cement and incubated overnight in a humid chamber at 37°C. After hybridization and post-

hybridization washes, the biotinylated probe is detected by avidin-FITC and amplified with additional layers of

5   biotinylated goat anti-avidin and avidin-FITC. For chromosomal localization, fluorescent R-bands are obtained as

previously described (Cherif et al., *supra*.). The slides are observed under a LEICA fluorescence microscope (DMRXA).

Chromosomes are counterstained with propidium iodide and the fluorescent signal of the probe appears as two

symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome (red). Thus, a particular

extended cDNA (or genomic DNA obtainable therefrom) may be localized to a particular cytogenetic R-band on a given

10 chromosome.

Once the extended cDNAs (or genomic DNAs obtainable therefrom) have been assigned to particular

chromosomes using the techniques described in Examples 49-51 above, they may be utilized to construct a high

resolution map of the chromosomes on which they are located or to identify the chromosomes in a sample.

EXAMPLE 52

15   Use of Extended cDNAs to Construct or Expand Chromosome Maps

Chromosome mapping involves assigning a given unique sequence to a particular chromosome as described

above. Once the unique sequence has been mapped to a given chromosome, it is ordered relative to other unique

sequences located on the same chromosome. One approach to chromosome mapping utilizes a series of yeast artificial

chromosomes (YACs) bearing several thousand long inserts derived from the chromosomes of the organism from which

20   the extended cDNAs (or genomic DNAs obtainable therefrom) are obtained. This approach is described in Ramaiah

Nagaraja et al. **Genome Research** 7:210-222, March 1997. Briefly, in this approach each chromosome is broken into

overlapping pieces which are inserted into the YAC vector. The YAC inserts are screened using PCR or other methods to

determine whether they include the extended cDNA (or genomic DNA obtainable therefrom) whose position is to be

determined. Once an insert has been found which includes the extended cDNA (or genomic DNA obtainable therefrom),

25   the insert can be analyzed by PCR or other methods to determine whether the insert also contains other sequences

known to be on the chromosome or in the region from which the extended cDNA (or genomic DNA obtainable therefrom)

was derived. This process can be repeated for each insert in the YAC library to determine the location of each of the

extended cDNAs (or genomic DNAs obtainable therefrom) relative to one another and to other known chromosomal

markers. In this way, a high resolution map of the distribution of numerous unique markers along each of the organisms

30   chromosomes may be obtained.

As described in Example 53 below extended cDNAs (or genomic DNAs obtainable therefrom) may also be used

to identify genes associated with a particular phenotype, such as hereditary disease or drug response.

EXAMPLE 53

Identification of genes associated with hereditary diseases or drug response

-77-

This example illustrates an approach useful for the association of extended cDNAs (or genomic DNAs obtainable therefrom) with particular phenotypic characteristics. In this example, a particular extended cDNA (or genomic DNA obtainable therefrom) is used as a test probe to associate that extended cDNA (or genomic DNA obtainable therefrom) with a particular phenotypic characteristic.

5   Extended cDNAs (or genomic DNAs obtainable therefrom) are mapped to a particular location on a human chromosome using techniques such as those described in Examples 49 and 50 or other techniques known in the art. A search of Mendelian Inheritance in Man (V. McKusick, **Mendelian Inheritance in Man** (available on line through Johns Hopkins University Welch Medical Library) reveals the region of the human chromosome which contains the extended cDNA (or genomic DNA obtainable therefrom) to be a very gene rich region containing several known genes and several

10   diseases or phenotypes for which genes have not been identified. The gene corresponding to this extended cDNA (or genomic DNA obtainable therefrom) thus becomes an immediate candidate for each of these genetic diseases.

Cells from patients with these diseases or phenotypes are isolated and expanded in culture. PCR primers from the extended cDNA (or genomic DNA obtainable therefrom) are used to screen genomic DNA, mRNA or cDNA obtained from the patients. Extended cDNAs (or genomic DNAs obtainable therefrom) that are not amplified in the patients can

15   be positively associated with a particular disease by further analysis. Alternatively, the PCR analysis may yield fragments of different lengths when the samples are derived from an individual having the phenotype associated with the disease than when the sample is derived from a healthy individual, indicating that the gene containing the extended cDNA may be responsible for the genetic disease.

VI. Use of Extended cDNAs (or genomic DNAs obtainable therefrom) to Construct Vectors

20   The present extended cDNAs (or genomic DNAs obtainable therefrom) may also be used to construct secretion vectors capable of directing the secretion of the proteins encoded by genes inserted in the vectors. Such secretion vectors may facilitate the purification or enrichment of the proteins encoded by genes inserted therein by reducing the number of background proteins from which the desired protein must be purified or enriched. Exemplary secretion vectors are described in Example 54 below.

25   **EXAMPLE 54**

                          Construction of Secretion Vectors

The secretion vectors of the present invention include a promoter capable of directing gene expression in the host cell, tissue, or organism of interest. Such promoters include the Rous Sarcoma Virus promoter, the SV40 promoter, the human cytomegalovirus promoter, and other promoters familiar to those skilled in the art.

30   A signal sequence from an extended cDNA (or genomic DNA obtainable therefrom), such as one of the signal sequences in SEQ ID NOs: 40-140 and 242-377 as defined in Table IV above, is operably linked to the promoter such that the mRNA transcribed from the promoter will direct the translation of the signal peptide. The host cell, tissue, or organism may be any cell, tissue, or organism which recognizes the signal peptide encoded by the signal sequence in the

-78-

extended cDNA (or genomic DNA obtainable therefrom). Suitable hosts include mammalian cells, tissues or organisms, avian cells, tissues, or organisms, insect cells, tissues or organisms, or yeast.

In addition, the secretion vector contains cloning sites for inserting genes encoding the proteins which are to be secreted. The cloning sites facilitate the cloning of the insert gene in frame with the signal sequence such that a fusion
5 protein in which the signal peptide is fused to the protein encoded by the inserted gene is expressed from the mRNA transcribed from the promoter. The signal peptide directs the extracellular secretion of the fusion protein.

The secretion vector may be DNA or RNA and may integrate into the chromosome of the host, be stably maintained as an extrachromosomal replicon in the host, be an artificial chromosome, or be transiently present in the host. Many nucleic acid backbones suitable for use as secretion vectors are known to those skilled in the art, including
10 retroviral vectors, SV40 vectors, Bovine Papilloma Virus vectors, yeast integrating plasmids, yeast episomal plasmids, yeast artificial chromosomes, human artificial chromosomes, P element vectors, baculovirus vectors, or bacterial plasmids capable of being transiently introduced into the host.

The secretion vector may also contain a polyA signal such that the polyA signal is located downstream of the gene inserted into the secretion vector.

15 After the gene encoding the protein for which secretion is desired is inserted into the secretion vector, the secretion vector is introduced into the host cell, tissue, or organism using calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection, viral particles or as naked DNA. The protein encoded by the inserted gene is then purified or enriched from the supernatant using conventional techniques such as ammonium sulfate precipitation, immunoprecipitation, immunochromatography, size exclusion chromatography, ion exchange
20 chromatography, and hplc. Alternatively, the secreted protein may be in a sufficiently enriched or pure state in the supernatant or growth media of the host to permit it to be used for its intended purpose without further enrichment.

The signal sequences may also be inserted into vectors designed for gene therapy. In such vectors, the signal sequence is operably linked to a promoter such that mRNA transcribed from the promoter encodes the signal peptide. A cloning site is located downstream of the signal sequence such that a gene encoding a protein whose secretion is
25 desired may readily be inserted into the vector and fused to the signal sequence. The vector is introduced into an appropriate host cell. The protein expressed from the promoter is secreted extracellularly, thereby producing a therapeutic effect.

The extended cDNAs or 5' ESTs may also be used to clone sequences located upstream of the extended cDNAs or 5' ESTs which are capable of regulating gene expression, including promoter sequences, enhancer sequences, and
30 other upstream sequences which influence transcription or translation levels. Once identified and cloned, these upstream regulatory sequences may be used in expression vectors designed to direct the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative fashion. Example 55 describes a method for cloning sequences upstream of the extended cDNAs or 5' ESTs.

EXAMPLE 55

Use of Extended cDNAs or 5' ESTs to Clone Upstream
Sequences from Genomic DNA

Sequences derived from extended cDNAs or 5' ESTs may be used to isolate the promoters of the corresponding genes using chromosome walking techniques. In one chromosome walking technique, which utilizes the

5 GenomeWalker™ kit available from Clontech, five complete genomic DNA samples are each digested with a different restriction enzyme which has a 6 base recognition site and leaves a blunt end. Following digestion, oligonucleotide adapters are ligated to each end of the resulting genomic DNA fragments.

For each of the five genomic DNA libraries, a first PCR reaction is performed according to the manufacturer's instructions using an outer adaptor primer provided in the kit and an outer gene specific primer. The gene specific primer

10 should be selected to be specific for the extended cDNA or 5' EST of interest and should have a melting temperature, length, and location in the extended cDNA or ' EST which is consistent with its use in PCR reactions. Each first PCR reaction contains 5ng of genomic DNA, 5 μl of 10X Tth reaction buffer, 0.2 mM of each dNTP, 0.2 μM each of outer adaptor primer and outer gene specific primer, 1.1 mM of Mg(OAc)$_2$, and 1 μl of the Tth polymerase 50X mix in a total volume of 50 μl. The reaction cycle for the first PCR reaction is as follows: 1 min @ 94°C / 2 sec @ 94°C, 3 min @

15 72°C (7 cycles) / 2 sec @ 94°C, 3 min @ 67°C (32 cycles) / 5 min @ 67°C.

The product of the first PCR reaction is diluted and used as a template for a second PCR reaction according to the manufacturer's instructions using a pair of nested primers which are located internally on the amplicon resulting from the first PCR reaction. For example, 5 μl of the reaction product of the first PCR reaction mixture may be diluted 180 times. Reactions are made in a 50 μl volume having a composition identical to that of the first PCR reaction except

20 the nested primers are used. The first nested primer is specific for the adaptor, and is provided with the GenomeWalker™ kit. The second nested primer is specific for the particular extended cDNA or 5' EST for which the promoter is to be cloned and should have a melting temperature, length, and location in the extended cDNA or 5' EST which is consistent with its use in PCR reactions. The reaction parameters of the second PCR reaction are as follows: 1 min @ 94°C / 2 sec @ 94°C, 3 min @ 72°C (6 cycles) / 2 sec @ 94°C, 3 min @ 67°C (25 cycles) / 5 min @ 67°C

25 The product of the second PCR reaction is purified, cloned, and sequenced using standard techniques. Alternatively, two or more human genomic DNA libraries can be constructed by using two or more restriction enzymes. The digested genomic DNA is cloned into vectors which can be converted into single stranded, circular, or linear DNA. A biotinylated oligonucleotide comprising at least 15 nucleotides from the extended cDNA or 5' EST sequence is hybridized to the single stranded DNA. Hybrids between the biotinylated oligonucleotide and the single stranded DNA containing

30 the extended cDNA or EST sequence are isolated as described in Example 29 above. Thereafter, the single stranded DNA containing the extended cDNA or EST sequence is released from the beads and converted into double stranded DNA using a primer specific for the extended cDNA or 5' EST sequence or a primer corresponding to a sequence included in the cloning vector. The resulting double stranded DNA is transformed into bacteria. DNAs containing the 5' EST or extended cDNA sequences are identified by colony PCR or colony hybridization.

-80-

Once the upstream genomic sequences have been cloned and sequenced as described above, prospective promoters and transcription start sites within the upstream sequences may be identified by comparing the sequences upstream of the extended cDNAs or 5' ESTs with databases containing known transcription start sites, transcription factor binding sites, or promoter sequences.

5   In addition, promoters in the upstream sequences may be identified using promoter reporter vectors as described in Example 56.

EXAMPLE 56

Identification of Promoters in Cloned Upstream Sequences

The genomic sequences upstream of the extended cDNAs or 5' ESTs are cloned into a suitable promoter

10 reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, β galactosidase, or green fluorescent protein. The sequences upstream of the extended cDNAs or 5' ESTs are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The

15 level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for augmenting transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the

20 inserted upstream sequence.

Appropriate host cells for the promoter reporter vectors may be chosen based on the results of the above described determination of expression patterns of the extended cDNAs and ESTs. For example, if the expression pattern analysis indicates that the mRNA corresponding to a particular extended cDNA or 5' EST is expressed in fibroblasts, the promoter reporter vector may be introduced into a human fibroblast cell line.

25 Promoter sequences within the upstream genomic DNA may be further defined by constructing nested deletions in the upstream DNA using conventional techniques such as Exonuclease III digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity. In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate

30 potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into the cloning sites in the promoter reporter vectors.

EXAMPLE 57

Cloning and Identification of Promoters

-81-

Using the method described in Example 55 above with 5' ESTs, sequences upstream of several genes were obtained. Using the primer pairs GGG AAG ATG GAG ATA GTA TTG CCT G (SEQ ID NO:29) and CTG CCA TGT ACA TGA TAG AGA GAT TC (SEQ ID NO:30), the promoter having the internal designation P13H2 (SEQ ID NO:31) was obtained.

5   Using the primer pairs GTA CCA GGGG ACT GTG ACC ATT GC (SEQ ID NO:32) and CTG TGA CCA TTG CTC CCA AGA GAG (SEQ ID NO:33), the promoter having the internal designation P15B4 (SEQ ID NO:34) was obtained.

Using the primer pairs CTG GGA TGG AAG GCA CGG TA (SEQ ID NO:35) and GAG ACC ACA CAG CTA GAC AA (SEQ ID NO:36), the promoter having the internal designation P2986 (SEQ ID NO:37) was obtained.

Figure 8 provides a schematic description of the promoters isolated and the way they are assembled with the
10   corresponding 5' tags. The upstream sequences were screened for the presence of motifs resembling transcription factor binding sites or known transcription start sites using the computer program MatInspector release 2.0, August 1996.

Figure 9 describes the transcription factor binding sites present in each of these promoters. The columns labeled matrice provides the name of the MatInspector matrix used. The column labeled position provides the 5' postion
15   of the promoter site. Numeration of the sequence starts from the transcription site as determined by matching the genomic sequence with the 5' EST sequence. The column labeled "orientation" indicates the DNA strand on which the site is found, with the + strand being the coding strand as determined by matching the genomic sequence with the sequence of the 5' EST. The column labeled "score" provides the MatInspector score found for this site. The column labeled "length" provides the length of the site in nucleotides. The column labeled "sequence" provides the sequence of
20   the site found.

The promoters and other regulatory sequences located upstream of the extended cDNAs or 5' ESTs may be used to design expression vectors capable of directing the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative manner. A promoter capable of directing the desired spatial, temporal, developmental, and quantitative patterns may be selected using the results of the expression analysis described in Example 26 above. For
25   example, if a promoter which confers a high level of expression in muscle is desired, the promoter sequence upstream of an extended cDNA or 5' EST derived from an mRNA which is expressed at a high level in muscle, as determined by the method of Example 26, may be used in the expression vector.

Preferably, the desired promoter is placed near multiple restriction sites to facilitate the cloning of the desired insert downstream of the promoter, such that the promoter is able to drive expression of the inserted gene. The
30   promoter may be inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration into the host chromosomes or transient expression. Suitable backbones for the present expression vectors include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

-82-

Preferably, the expression vectors also include a polyA signal downstream of the multiple restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

Following the identification of promoter sequences using the procedures of Examples 55-57, proteins which interact with the promoter may be identified as described in Example 58 below.

5   EXAMPLE 58

<u>Identification of Proteins Which Interact with Promoter Sequences, Upstream</u>

<u>Regulatory Sequences, or mRNA</u>

Sequences within the promoter region which are likely to bind transcription factors may be identified by homology to known transcription factor binding sites or through conventional mutagenesis or deletion analyses of

10   reporter plasmids containing the promoter sequence. For example, deletions may be made in a reporter plasmid containing the promoter sequence of interest operably linked to an assayable reporter gene. The reporter plasmids carrying various deletions within the promoter region are transfected into an appropriate host cell and the effects of the deletions on expression levels is assessed. Transcription factor binding sites within the regions in which deletions reduce expression levels may be further localized using site directed mutagenesis, linker scanning analysis, or other

15   techniques familiar to those skilled in the art. Nucleic acids encoding proteins which interact with sequences in the promoter may be identified using one-hybrid systems such as those described in the manual accompanying the Matchmaker One-Hybrid System kit avalilabe from Clontech (Catalog No. K1603-1). Briefly, the Matchmaker One-hybrid system is used as follows. The target sequence for which it is desired to identify binding proteins is cloned upstream of a selectable reporter gene and integrated into the yeast genome. Preferably, multiple copies of the target sequences are

20   inserted into the reporter plasmid in tandem.

A library comprised of fusions between cDNAs to be evaluated for the ability to bind to the promoter and the activation domain of a yeast transcription factor, such as GAL4, is transformed into the yeast strain containing the integrated reporter sequence. The yeast are plated on selective media to select cells expressing the selectable marker linked to the promoter sequence. The colonies which grow on the selective media contain genes encoding proteins which

25   bind the target sequence. The inserts in the genes encoding the fusion proteins are further characterized by sequencing. In addition, the inserts may be inserted into expression vectors or in vitro transcription vectors. Binding of the polypeptides encoded by the inserts to the promoter DNA may be confirmed by techniques familiar to those skilled in the art, such as gel shift analysis or DNAse protection analysis.

VII. Use of Extended cDNAs (or Genomic DNAs Obtainable Therefrom) in Gene Therapy

30   The present invention also comprises the use of extended cDNAs (or genomic DNAs obtainable therefrom) in gene therapy strategies, including antisense and triple helix strategies as described in Examples 57 and 58 below. In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the expression of the protein encoded by the mRNA. The antisense sequences may prevent gene expression through a variety of mechanisms. For example, the antisense sequences may inhibit the ability of ribosomes

-83-

to translate the mRNA. Alternatively, the antisense sequences may block transport of the mRNA from the nucleus to the cytoplasm, thereby limiting the amount of mRNA available for translation. Another mechanism through which antisense sequences may inhibit gene expression is by interfering with mRNA splicing. In yet another strategy, the antisense nucleic acid may be incorporated in a ribozyme capable of specifically cleaving the target mRNA.

5   **EXAMPLE 59**

                               Preparation and Use of Antisense Oligonucleotides

        The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. They may comprise a sequence complementary to the sequence of the extended cDNA (or genomic DNA obtainable therefrom). The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an

10   intracellular duplex having sufficient stability to inhibit the expression of the mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green et al., **Ann. Rev. Biochem.** 55:569-597 (1986) and Izant and Weintraub, **Cell** 36:1007-1015 (1984).

        In some strategies, antisense molecules are obtained from a nucleotide sequence encoding a protein by reversing the orientation of the coding region with respect to a promoter so as to transcribe the opposite strand from

15   that which is normally transcribed in the cell. The antisense molecules may be transcribed using in vitro transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of the antisense nucleic acids in vivo by operably linking DNA containing the antisense sequence to a promoter in an expression vector.

        Alternatively, oligonucleotides which are complementary to the strand normally transcribed in the cell may be

20   synthesized in vitro. Thus, the antisense nucleic acids are complementary to the corresponding mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of modifications suitable for use in antisense strategies are described by Rossi et al., **Pharmacol. Ther.** 50(2):245-254, (1991).

25   Various types of antisense oligonucleotides complementary to the sequence of the extended cDNA (or genomic DNA obtainable therefrom) may be used. In one preferred embodiment, stable and semi-stable antisense oligonucleotides described in International Application No. PCT WO94/23026 are used. In these moleucles, the 3' end or both the 3' and 5' ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are better able to withstand exonuclease attacks and exhibit increased stability compared to conventional antisense

30   oligonucleotides.

        In another preferred embodiment, the antisense oligodeoxynucleotides against herpes simplex virus types 1 and 2 described in International Application No. WO 95/04141.

        In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523 are used. These double- or single-stranded oligonucleotides comprise one or

-84-

more, respectively, inter- or intra-oligonucleotide covalent cross-linkages, wherein the linkage consists of an amide bond

between a primary amine group of one strand and a carboxyl group of the other strand or of the same strand,

respectively, the primary amine group being directly substituted in the 2' position of the strand nucleotide

monosaccharide ring, and the carboxyl group being carried by an aliphatic spacer group substituted on a nucleotide or

5   nucleotide analog of the other strand or the same strand, respectively.

The antisense oligodeoxynucleotides and oligonucleotides disclosed in International Application No. WO

92/18522 may also be used. These molecules are stable to degradation and contain at least one transcription control

recognition sequence which binds to control proteins and are effective as decoys therefor. These molecules may contain

"hairpin" structures, "dumbbell" structures, "modified dumbbell" structures, "cross-linked" decoy structures and "loop"

10   structures.

In another preferred embodiment, the cyclic double-stranded oligonucleotides described in European Patent

Application No. 0 572 287 A2 are used. These ligated oligonucleotide "dumbbells" contain the binding site for a

transcription factor and inhibit expression of the gene under control of the transcription factor by sequestering the

factor.

15   Use of the closed antisense oligonucleotides disclosed in International Application No. WO 92/19732 is also

contemplated. Because these molecules have no free ends, they are more resistant to degradation by exonucleases than

are conventional oligonucleotides. These oligonucleotides may be multifunctional, interacting with several regions which

are not adjacent to the target mRNA.

The appropriate level of antisense nucleic acids required to inhibit gene expression may be determined using in

20   vitro expression analysis. The antisense molecule may be introduced into the cells by diffusion, injection, infection or

transfection using procedures known in the art. For example, the antisense nucleic acids can be introduced into the body

as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral

protein, or as an oligonucleotide operably linked to a promoter contained in an expression vector. The expression vector

may be any of a variety of expression vectors known in the art, including retroviral or viral vectors, vectors capable of

25   extrachromosomal replication, or integrating vectors. The vectors may be DNA or RNA.

The antisense molecules are introduced onto cell samples at a number of different concentrations preferably

between 1×10^{-10}M to 1×10^{-4}M. Once the minimum concentration that can adequately control gene expression is

identified, the optimized dose is translated into a dosage suitable for use in vivo. For example, an inhibiting

concentration in culture of 1×10^{-7} translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of

30   oligonucleotide approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the

oligonucleotide in laboratory animals. It is additionally contemplated that cells from the vertebrate are removed, treated

with the antisense oligonucleotide, and reintroduced into the vertebrate.

-85-

It is further contemplated that the antisense oligonucleotide sequence is incorporated into a ribozyme sequence to enable the antisense to specifically bind and cleave its target mRNA. For technical applications of ribozyme and antisense oligonucleotides see Rossi et al., *supra*.

In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that the
5   effectiveness of antisense inhibition on translation can be monitored using techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabeling.

The extended cDNAs of the present invention (or genomic DNAs obtainable therefrom) may also be used in gene therapy approaches based on intracellular triple helix formation. Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity as it is associated with
10   a particular gene. The extended cDNAs (or genomic DNAs obtainable therefrom) of the present invention or, more preferably, a portion of those sequences, can be used to inhibit gene expression in individuals having diseases associated with expression of a particular gene. Similarly, a portion of the extended cDNA (or genomic DNA obtainable therefrom) can be used to study the effect of inhibiting transcription of a particular gene within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also
15   inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the extended cDNA or from the gene corresponding to the extended cDNA are contemplated within the scope of this invention.

EXAMPLE 60

Preparation and use of Triple Helix Probes

20   The sequences of the extended cDNAs (or genomic DNAs obtainable therefrom) are scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting gene expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting gene expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which normally express the target gene. The oligonucleotides may be prepared on an oligonucleotide
25   synthesizer or they may be purchased commercially from a company specializing in custom oligonucleotide synthesis, such as GENSET, Paris, France.

The oligonucleotides may be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

30   Treated cells are monitored for altered cell function or reduced gene expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the target gene in cells which have been treated with the oligonucleotide . The cell functions to be monitored are predicted based upon the homologies of the target gene corresponding to the extended cDNA from which the oligonucleotide was derived with known gene sequences that have been associated with a particular function. The cell functions can also be

-86-

predicted based on the presence of abnormal physiologies within cells derived from individuals with a particular inherited disease, particularly when the extended cDNA is associated with the disease using techniques described in Example 53.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced in vivo using the techniques described above and in Example 59 at a dosage calculated based on the in vitro

5 results, as described in Example 59.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin et al. (Science 245:967-

10 971 (1989).

EXAMPLE 61

Use of Extended cDNAs to Express an Encoded Protein in a Host Organism

The extended cDNAs of the present invention may also be used to express an encoded protein in a host organism to produce a beneficial effect. In such procedures, the encoded protein may be transiently expressed in the

15 host organism or stably expressed in the host organism. The encoded protein may have any of the activities described above. The encoded protein may be a protein which the host organism lacks or, alternatively, the encoded protein may augment the existing levels of the protein in the host organism.

A full length extended cDNA encoding the signal peptide and the mature protein, or an extended cDNA encoding only the mature protein is introduced into the host organism. The extended cDNA may be introduced into the host

20 organism using a variety of techniques known to those of skill in the art. For example, the extended cDNA may be injected into the host organism as naked DNA such that the encoded protein is expressed in the host organism, thereby producing a beneficial effect.

Alternatively, the extended cDNA may be cloned into an expression vector downstream of a promoter which is active in the host organism. The expression vector may be any of the expression vectors designed for use in gene

25 therapy, including viral or retroviral vectors.

The expression vector may be directly introduced into the host organism such that the encoded protein is expressed in the host organism to produce a beneficial effect. In another approach, the expression vector may be introduced into cells in vitro. Cells containing the expression vector are thereafter selected and introduced into the host organism, where they express the encoded protein to produce a beneficial effect.

30 ## EXAMPLE 62

Use Of Signal Peptides Encoded By 5' Ests Or Sequences

Obtained Therefrom To Import Proteins Into Cells

The short core hydrophobic region (h) of signal peptides encoded by the 5'ESTS or extended cDNAs derived from the 5'ESTs of the present invention may also be used as a carrier to import a peptide or a protein of interest, so-

called cargo, into tissue culture cells (Lin *et al.*, *J. Biol. Chem.*, 270: 14225-14258 (1995); Du *et al.*, *J. Peptide Res.*, 51: 235-243 (1998); Rojas *et al.*, *Nature Biotech.*, 16: 370-375 (1998)).

When cell permeable peptides of limited size (approximately up to 25 amino acids) are to be translocated across cell membrane, chemical synthesis may be used in order to add the h region to either the C-terminus or the N-

5 terminus to the cargo peptide of interest. Alternatively, when longer peptides or proteins are to be imported into cells, nucleic acids can be genetically engineered, using techniques familiar to those skilled in the art, in order to link the extended cDNA sequence encoding the h region to the 5' or the 3' end of a DNA sequence coding for a cargo polypeptide. Such genetically engineered nucleic acids are then translated either *in vitro* or *in vivo* after transfection into appropriate cells, using conventional techniques to produce the resulting cell permeable polypeptide. Suitable hosts cells are then

10 simply incubated with the cell permeable polypeptide which is then translocated across the membrane.

This method may be applied to study diverse intracellular functions and cellular processes. For instance, it has been used to probe functionally relevant domains of intracellular proteins and to examine protein-protein interactions involved in signal transduction pathways (Lin *et al.*, *supra*; Lin *et al.*, *J. Biol. Chem.*, 271: 5305-5308 (1996); Rojas *et al.*, *J. Biol. Chem.*, 271: 27456-27461 (1996); Liu *et al.*, *Proc. Natl. Acad. Sci. USA*, 93: 11819-11824 (1996); Rojas *et*

15 *al.*, *Bioch. Biophys. Res. Commun.*, 234: 675-680 (1997)).

Such techniques may be used in cellular therapy to import proteins producing therapeutic effects. For instance, cells isolated from a patient may be treated with imported therapeutic proteins and then re-introduced into the host organism.

Alternatively, the h region of signal peptides of the present invention could be used in combination with a

20 nuclear localization signal to deliver nucleic acids into cell nucleus. Such oligonucleotides may be antisense oligonucleotides or oligonucleotides designed to form triple helixes, as described in examples 59 and 60 respectively, in order to inhibit processing and maturation of a target cellular RNA.

EXAMPLE 63

Reassembling & Resequencing of Clones

25 Full length cDNA clones obtained by the procedure described in Example 27 were double-sequenced. These sequences were assembled and the resulting consensus sequences were then reanalyzed. Open reading frames were reassigned following essentially the same process as the one described in Example 27.

After this reanalysis process a few abnormalities were revealed. The sequences presented in SEQ ID NOs: 47, 73, 79, 89, 91, 96, 126, 128, 134, and 139 are apparently unlikely to be genuine full length cDNAs. These clones are

30 missing a stop codon and are thus more probably 3' truncated cDNA sequences. Similarly, the sequences presented in SEQ ID NOs: 45, 50, 54, 57, 73, 74, 89, 92, 95, 98, 126, 129, 130, 131 and 139 may also not be genuine full length cDNAs based on homology studies with existing protein sequences. Although both of these sequences encode a potential start methionine each could represent a 5' truncated cDNA.

-88-

In addition, SEQ ID NO: 115 was found to be an alternatively spliced transcript and the identities of some of the bases in SEQ ID NO: 131 were corrected.

Finallly, after the reassignment of open reading frames for the clones, new open reading frames were chosen in some instances. For example, in the case of SEQ ID NOs: 41, 47, 50, 52, 54-56, 58, 59, 61, 74, 75, 79, 84, 89, 91, 92,
5   96, 98, 103, 105, 106, 126, 129, 131, and 133 the new open reading frames were no longer predicted to contain a signal peptide.

As discussed above, Table IV provides the sequence identification numbers of the extended cDNAs of the present invention, the locations of the full coding sequences in SEQ ID NOs: 40-140 and 242-377 (i.e. the nucleotides encoding both the signal peptide and the mature protein, listed under the heading FCS location in Table IV), the locations
10   of the nucleotides in SEQ ID NOs: 40-140 and 242-377 which encode the signal peptides (listed under the heading SigPep Location in Table IV), the locations of the nucleotides in SEQ ID NOs: 40-140 and 242-377 which encode the mature proteins generated by cleavage of the signal peptides (listed under the heading Mature Polypeptide Location in Table IV), the locations in SEQ ID NOs: 40-140 and 242-377 of stop codons (listed under the heading Stop Codon Location in Table IV) the locations in SEQ ID NOs: 40-140 and 242-377 of polyA signals (listed under the heading g
15   PolyA Signal Location in Table IV) and the locations of polyA sites (listed under the heading PolyA Site Location in Table IV).

As discussed above, Table V lists the sequence identification numbers of the polypeptides of SEQ ID NOs: 141-241 and 378-513, the locations of the amino acid residues of SEQ ID NOs: 141-241 and 378-513 in the full length polypeptide (second column), the locations of the amino acid residues of SEQ ID NOs: 141-241 and 378-513 in the signal
20   peptides (third column), and the locations of the amino acid residues of SEQ ID NOs: 141-241 and 379-513 in the mature polypeptide created by cleaving the signal peptide from the fall length polypeptide (fourth column). In Table V, and in the appended sequence listing, the first amino acid of the mature protein resulting from cleavage of the signal peptide is designated as amino acid number 1 and the first amino acid of the signal peptide is designated with the appropriate negative number, in accordance with the regulations governing sequence listings.

25   **EXAMPLE 64**

                            Functional Analysis of Predicted Protein Sequences

Following double-sequencing, new contigs were assembled for each of the extended cDNAs of the present invention and each was compared to known sequences available at the time of filing. These sequences originate from the following databases: Genbank (release 108 and daily releases up to October, 15, 1998), Genseq (release 32) PIR (release
30   53) and SwissProt (release 35). The predicted proteins of the present invention matching known proteins were further classified into 3 categories depending on the level of homology.

The first category contains proteins of the present invention exhibiting more than 70% identical amino acid residues on the whole length of the matched protein. They are clearly close homologues which most probably have the same function or a very similar function as the matched protein.

-89-

The second category contains proteins of the present invention exhibiting more remote homologies (40 to 70% over the whole protein) indicating that the protein of the present inventionmay have functions similar to those of the homologous protein.

The third category contains proteins exhibiting homology (90 to 100%) to a domain of a known protein
5 indicating that the matched protein and the protein of the invention may share similar features.

It should be noted that the numbering of amino acids in the protein sequences discussed in Figures 10 to 15, and Table VIII, the first methionine encountered is designated as amino acid number 1. In the appended sequence listing, the first amino acid of the mature protein resulting from cleavage of the signal peptide is designated as amino acid · number 1, and the first amino acid of the signal peptide is designated with the appropriate negative number, in
10 accordance with the regulations governing sequence listings.

In addition all of the corrected amino acid sequences (SEQ ED NOs: 141-241 and 378-513) were scanned for the presence of known protein signatures and motifs. This search was performed agains\ the Prosite 15.0 database, using the Proscan software from the GCG package- Functional signatures and their locations are indicated in Table VIII.

15 **A) Proteins which are closely related to known proteins**

<u>Protein of SEQ ID NO: 217</u>

. .:The protein of SEQ ID NO: 217 encoded by the extended cDNA SEQ ID NO: 116 isolated from lymphocyte shows complete identity to a human protein TFAR19 that may play a role in apoptosis (Genbank accession number AF014955, SEQ ID NO: 516) as shown by the alignment in figure 10.

20   Taken together, these data suggest that the protein of SEQ ID NO: 217 may be involved in the control of development and homeostasis. Thus, this protein may be useful in diagnosis and/or treating several types of disorders including, but not limited to, cancer, autoimmune disorders, viral infections such as AIDS, neurodegenerative disorders, osteoporosis.

25 <u>Proteins of SEQ ID NOs: 174, 175 and 232</u>

The proteins of SEQ ID NOs: 174, 175 and 232 encoded by the extended cDNAs SEQ ID NOs:. 73, 74 and 131 respectively and isolated from lymphocytes shows complete extensive homologies to a human secreted protein (Genseq accession number W36955, SEQ ID NO: 517). As shown by the alignments of figure 11, the amino acid residues are identical to those of the 110 amino acid long matched protein except for positions 51 and 108-110 of the matched
30 protein for the protein of SEQ ID NOs: 174, for positions 48, 94 and 108-110 of the matched protein of SEQ ID NOs:175 and for positions 94, and 108-110 of the matched protein for the protein of SEQ ID NOs: 232. Proteins of SEQ ID NOs: 174 and 232 may represent alternative forms issued from alternative use of polyadenylation signals.

Taken together, these data suggest that the proteins of SEQ ID NOs: 174. 175 and 232 may play a role in cell proliferation and/or differentiation, in immune responses and/or in haematopoesis. Thus. this protein or part therein,

may be useful in diagnosing and treating several disorders including, but not limited to, cancer, immunological,
haematological and/or inflammatory disorders. It may also be useful in modulating the immune and inflammatory
responses to infectious agents and/or to suppress graft rejection .

5   Proteins of SEQ ID NO: 231

     The protein of SEQ ID NO: 231 encoded by the extended cDNA SEQ ID NO: 130 shows extensive homology
with the human E25 protein (Genbank accession number AF038953, SEQ ID NO: 515). As shown by the alignments in
figure 12, the amino acid residues are identical except for position 159 in the 263 amino acid long matched sequence.
The matched protein might be involved in the development and differentiation of haematopoietic stem/progenitor cells.
10 In addition, it is the human homologue of a murine protein thought to be involved in chondro-osteogenic differentiation
and belonging to a novel multigene family of integral membrane proteins (Deleersnijder *et al*, *J. Biol. Chem.*, 271 :
19475-19482 (1996)).

     The protein of invention contains two short segments from positions 1 to 21 and from 100 to 120 as predicted
by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10 : 685-686 (1994)). The first
15 transmembrane domains matches exactly those predicted for the murine E25 protein.

     Taken together, these data suggest that the protein of SEQ ID NO: 231 may be involved in cellular proliferation
and differentiation. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including,
but not limited to, cancer and embryogenesis disorders.

20 Protein of SEQ ID NO: 196

     The protein of SEQ ID NO: 196 encoded by the extended cDNA SEQ ID NO: 95 shows extensive homology with
the human seventransmembrane protein (Genbank accession number Y11395, SEQ ID NO: 518) and its murine
homologue (Genbank accession number Y11550). As shown by the alignments in figure 13, the amino acid residues are
identical except for position 174 in the 399 amino acid long human matched sequence. The matched protein potentially
25 associated to stomatin may act as a G-protein coupled receptor and is likely to be important for the signal transduction
in neurons and haematopoietic cells (Mayer *et al*, *Biochem. Biophys. Acta.*, 1395 : 301-308 (1998)).

     Taken together, these data suggest that the protein of SEQ ID NOs: 196 may be involved in signal
transduction. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not
limited to, cancer, neurodegenerative diseases cardiovascular disorders, hypertension, renal injury and repair and septic
30 shock.

Protein of SEQ ID NO: 158

     The protein of SEQ ID NOs: 158 encoded by the extended cDNA SEQ ID NO: 57 shows homology with the
murine subunit 7a of the COP9 complex (Genbank accession number AF071316, SEQ ID NO: 520). As shown by the

-91-

alignments in figure 14, the amino acid residues are identical except for positions 90, 172 and 247 in the 275 amino acid long matched sequence. This complex is highly conserved between mammals and higher plants where it has been shown to act as a repressor of photomorphogenesis All the components of the mammalian COP9 complex contain structural features also present in components of the proteasome regulatory complex and the translation initiation

5   complex eIF3 complex, suggesting that the mammalian COP9 complex is an important cellular regulator modulating multiple signaling pathways (Wei *et al, Curr. Biol.*, 8 : 919-922 (1998)).

Taken together, these data suggest that the protein of SEQ ID NO: 158 may be involved in cellular signaling, probably as a subunit of the human COP9 complex. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not limited to, cancer, neurodegenerative diseases, cardiovascular disorders,

10 hypertension, renal injury and repair and septic shock.

Protein of SEQ ID NO: 226

The protein of SEQ ID NO: 226 encoded by the extended cDNA SEQ ID NO: 125 shows homology with the bovine subunit B14.5B of the NADH-ubiquinone oxidureductase complex (Arizmendi *et al, FEBS Lett.*, 313 : 80-84

15 (1992) and Swissprot accession -number Q02827, SEQ ID NO: 514). As shown by the alignments in figure 15, the amino acid residues are identical except for positions 3-4, 6-12, 32-34, 47, 53-55, 67 and 69-74 in the 120 amino acid long matched sequence. This complex is the first of four complexes located in the inner mitochondrial membrane and composing the mitochondrial electron transport chain. Complex I is involved in the dehydrogenation of NADH and the transportation of electrons to coenzyme Q. It is composed of 7 subunits encoded by the mitochondrial genome and 34

20 subunits encoded by the nuclear genome. It is also thought to play a role in the regulation of apoptosis and necrosis. Mitochondriocytopathies due to complex I deficiency are frequently encountered and affect tissues with a high energy demand such as brain (mental retardation, convulsions, movement disorders), heart (cardiomyopathy, conduction disorders), kidney (Fanconi syndrome), skeletal muscle (exercise intolerance, muscle weakness, hypotonia) and/or eye (opthmaloplegia, ptosis, cataract and retinopathy). For a review on complex I see Smeitink *et al., Hum. Mol. Gent., 7* :

25 1573-1579 (1998).

Taken together, these data suggest that the protein of SEQ ID NO: 226 may be part of the mitochondrial energy-generating system, probably as a subunit of the NADH-ubiquinone oxidoreductase complex. Thus, this protein or part therein, may be useful in diagnosing and/or treating several disorders including, but not limited to, brain disorders (mental retardation, convulsions, movement disorders), heart disorders (cardiomyopathy, conduction disorders), kidney

30 disorders (Fanconi syndrome), skeletal muscle disorders (exercise intolerance, muscle weakness, hypotonia) and/or eye disorders opthmalmoplegia, ptosis, cataract and retinopathy).

B) Proteins which are remotely related to proteins with known functions
Proteins of SEQ ID NOs: 149, 150 and 211

·92·

The proteins of SEQ ID NOs: 1,49,150 and 211 encoded by the extended cDNAs SEQ ID NOs: 48, 49 and 110 respectively and found in, skeletal muscle shows homologies with T1/ST2 ligand polypeptide of either human (Genbank accession number U41804 and Genseq accession number W09639) or rodent species (Genbank accession number U41805 and Genseq accession number W09640). These polypeptides are thought to be cytokines that bind to the ST2

5   receptor, a member of the immunoglobulin family homologous to the interleukin-1 receptor and present on some lymphoma cells. They are predicted to be cell-surface proteins containing a short transmembrane domain. (Gayle *et al, J. Biol. Chem.*, 271 : 5784-5789 (1996)). Proteins of SEQ ID NOs: 149, 150 and 211 may represent alternative forms issued from alternative use of polyadenylation signals.

The protein of invention contains two short transmembrane segments from positions 5 to 25 and from 195 to

10   215 as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, **10** :685-686 (1994)). The second transmembrane domain matches exactly those of the matched cell-surface protein.

Taken together, these data suggest that the protein of SEQ ID NOs: 149, 150 and 211 may act as a cytokine, thus may play a role in the regulation of cell growth and differentiation and/or in the regulation of the immune response. Thus, this protein or part therein, may be useful in diagnosing and treating several disorders including, but not limited to,

15   cancer, immunological, haematological and/or inflammatory disorders. It may also be useful in modulating the immune and inflammatory responses to infectious agents such as HIV and/or to suppress graft rejection.

Protein of SEQ ID NO: 177

The protein SEQ ID NO: 177 found in testis encoded by the extended cDNA SEQ ID NO: 76 shows homologies

20   to serine protease inhibitor proteins belonging to the pancreatic trypsin inhibitor family (Kunitz) such as the extracellular proteinase inhibitor named chelonianin (Swissprot accession number P00993). The characteristic PROSITE signature of this family is conserved in the protein of the invention (positions 69 to 87) except for a drastic change of the last cysteine residue into an arginine residue.

Taken together, these data suggest that the protein of SEQ ID NO: 177 may be a protease inhibitor, probably

25   of the Kunitz family. Thus, this protein or part therein, may be useful in diagnosing and treating several disorders including but not limited to, cancer and neurodegenerative disorders such as Alzheimer's disease.

Protein of SEQ ID NO: 146

The protein SEQ ID NO: 146 encoded by the extended cDNA SEQ ID NO: 45 shows homology to human

30   apolipoprotein L (Genbank accession number AF019225). The matched protein is a secreted high density lipoprotein associated with apoA-I-containing lipoproteins which play a key role in reverse cholesterol transport.

Taken together, these data suggest that the protein of SEQ ID NO 146 may play a role in lipid metabolism. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not limited to,

-93-

hyperlipidemia, hypercholesterolemia, atherosclerosis, cardiovascular disorders such as, coronary heart disease, and
neurodegenerative disorders such as Alzheimer's disease or dementia.

Protein of SEQ ID NO: 163

5   The protein SEQ ED NO: 163 encoded by the extended cDNA SEQ ID NO: 62 shows homology to the yeast
autophagocytosis protein AUT1 (SwissProt accession number P40344). The matched protein is required for starvation-
induced non-specific bulk transport of cytoplasmic proteins to the vacuole.

Taken together, these data suggest that the protein of SEQ ID NO: 163 may play a role in protein transport.
Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not limited to,
10   autoimmune disorders and immune disorders due to dysfunction of antigen presentation.

C) Proteins homologous to a domain of a protein with known function
Protein of SEQ ID NO: 214

The protein of SEQ ID NO: 214 encoded by the extended cDNA SEQ ID NO: 113 and expressed in adult brain
15   shows extensive homology to part of the murine SHYC protein (Genbank accession number AF072697) which is
expressed in the developing and embryonic nervous system as well as along the olfactory pathway in adult brains
(Köster et al., Neuroscience Letters., 252 : 69-71 (1998)).

Taken together, these data suggest that the protein of SEQ ID NO: 214 may play a role in nervous system
development and function. Thus, this protein may be useful in diagnosing and/or treating cancer and/or brain disorders,
20   including neurodegenerative disorders such as Alzheimer's and Parkinson's diseases.

Protein of SEQ ID NO: 225

The protein of SEQ ID NO: 225 encoded by the extended cDNA SEQ ID NO: 124 and expressed in adult
prostate belong to the phosphatidylethanolainin-binding protein from which it exhibits the characteristic PROSITE
25   signature from positions 90 to 112 (see table VIII). Proteins from this widespread family, from nematodes to fly, yeast,
rodent and primate species, bind hydrophobic ligands such as phospholipids and nucleotides. They are mostly expressed
in brain and in testis and are thought to play a role in cell growth and/or maturation, in regulation of the sperm
maturation, motility and 'in membrane remodeling. They may act either through signal transduction or through
oxidoreduction reactions (for a review see Schoentgen and Jollès, FEBS Letters, 369 : 22-26 (1995)).

30   Taken together, these data suggest that the protein of SEQ ID NO: 225 may play a role in cell. Thus, these
growth, maturation and in membrane remodeling and/or may be related to male fertility. Thus, this protein may be useful
in diagnosing and/or treating cancer, neurodegenerative diseases, and/of, disorders related to male fertility and sterility.

Protein of SEQ ID NO: 153

The protein of SEQ ID NO: 153 encoded by the extended cDNA SEQ ID NO. 52 and expressed in brain exhibits homology to different integral membrane proteins. These membrane proteins include the nematode protein SRE-2 (Swissprot accession number O09273) that belongs to the multigene SRE family of *C. elegans* receptor-like proteins and a family of tricarboxylate carriers conserved between flies and mammals. One member of this matched family is the rat

5   tricarboxylate carrier (Genbank accession number S70011), an anion transporter localized in the inner membrane of mitochondria and involved in the biosynthesis of fatty acids and cholesterol. The protein of the invention contains a short transmembrane segments from positions 5 to 25 as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10:685-686 (1994)).

Taken together, these data suggest that the protein of SEQ ID NO: 153 may play a role in signal transduction

10 and/or in molecule transport. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not limited to, cancer, neurodegenerative diseases, immune disorders, cardiovascular disorders, hypertension, renal injury and repair and septic shock

Protein of SEQ ID NO: 213

15 The protein of SEQ ID NO: 213 encoded by the extended cDNA SEQ ID NO: 112 and expressed in brain exhibits homology with part of the tRNA pseudouridine 55 synthase found in *Escherichia Coli* (Swissprot accession number P09171). This bacterial protein belongs to the NAP57/CBF5/TRUB family of nucleolar proteins found in bacteria, yeasts and mammals involved in rRNA or tRNA biosynthesis, ribosomal subunit assembly and/or centromere/mircotubule binding.

20 .   Taken together, these data suggest that the protein of SEQ ID NO: 213 may play a role in rRNA or tRNA biogensis and function. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not limited to, cancer, hearing loss or disorders linked to chromosomal instability such as dyskeratosis.

Protein of SEQ ED NO: 240

25 The protein of SEQ ID NO: 240 encoded by the extended cDNA SEQ ID NO: 139 and expressed in brain exhibits homology with a family of eukaryotic cell surface antigens containing 4 transmembrane domains. The PROSITE signature for this family is conserved in the protein of the invention except for a substitution of an alanine residue in place of any of the following hydrophic residues : leucine, valine, isoleucine or methionine (positions 21 to 36).

The protein of the invention contains three short transmembrane segments from positions 6 to 26, 32 to 52

30 and from 56 to 76 as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10 : 685-686 (1994)). These transmembrane domains match the last three transmembrane domains of the matched protein family.

Taken together, these data suggest that the protein of SEQ ID NO: 240 may play a role in immunological and/or inflammatory responses, probably as a cell surface antigen. Thus, this protein or part therein, may be useful in diagnosing and treating several disorders including, but not limited to, cancer, immunological, haematological and/or

-95-

inflammatory disorders. It may also be useful in modulating the immune and inflammatory responses to infectious agents and/or to suppress graft rejection.

Protein of SEQ ID NO: 239

5   The protein of SEQ ID NO: 239 encoded by the extended cDNA SEQ ID NO: 138 exhibits homology with a conserved region in a family of NA+/H+ exchanger conserved in yeast, nematode and mammals. These cation/proton exchangers are integral membrane proteins with 5 transmembrane segments involved in intracellular pH regulation, maintenance of cell volume, reabsorption of sodium across specialized epithelia, vectorial transport and are also thought to play a role in signal transduction and especially in the induction of cell proliferation and in the induction of apoptosis.

10   The protein of invention contains four short transmembrane segments from positions 21 to 41, 48 to 68 and from 131 to 151 as predicted by the software TopPred II (Claros and von Heijne, *CABIOS applic. Notes*, 10 : 685-686 (1994)). The third and fourth transmembrane domains match the fourth and fifth transmembrane segments of the matched family of proteins.

         Taken together, these data suggest that the protein of SEQ ID NO: 239 may play a role in membrane

15 permeability and/or in signal transduction. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not limited to, cancer, neurodegenerative diseases, cardiovascular disorders, hypertension, renal injury and repair, septic shock as well as disorders of membrane permeability such as diarrhea.

Protein of SEQ ID NO: 200

20   The protein of SEQ ID NO: 200 encoded by the extended cDNA SEQ ED NO: 99 and expressed in brain exhibits extensive homology to the N-terminus of cell division cycle protein 23 (Genbank accession number AF053977) and also to a lesser extent to its homologue in *Saccharomyces cerevisiae*. The matched protein is required for chromosome segregation and is part of the anaphae-promoting complex necessary for cell cycle progression to mitosis.

         Taken together, these data suggest that the protein of SEQ ID NO: 200 may play a role in cellular mitosis.

25 Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not limited to, cancer and leukemia.

Protein of SEQ ID NO: 230

         The protein of SEQ ID NO: 230 encoded by the extended cDNA SEQ ID NO: 129 exhibits extensive homology to

30 the C-terminus of the eta subunit of T-complex polypeptide 1 conserved from yeasts to mammals, and even complete identity with the last 54 amino acid residues of the human protein (Genbank accession number AF026292). The matched protein is a chaperonin which assists the folding of actins and tubulins in eukaryotic cells upon ATP hydrolysis.

         Taken together, these data suggest that the protein of SEQ ID NO: 230 may play a role in the folding, transport, assembly and degradation of proteins. Thus, this protein may be useful in diagnosing and/or treating several

types of disorders including, but not limited to, cancer, cardiovascular disorders, immune disorders, neurodegenerative

disorders, osteoporosis and arthritis.

Protein of SEQ ED NO: 167

5   The protein of SEQ ID NO: 167 encoded by the extended cDNA SEQ ID NO: 66 exhibits homology to a monkey

pepsinogen A-4 precursor (Swissprot accession number P27678) and to related members of the aspartyl protease family.

The matched protein belongs to a family of widely distributed proteolytic enzymes known to exist in vertebrate, fungi,

plants, retroviruses and some plant viruses.

           Taken together, these data suggest that the protein of SEQ ID NO: 167 may play a role in the degradation of

10   proteins. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not

limited to, cancer, autoimmune disorders and immune disorders due to dysfunction of antigen presentation.

Protein of SEQ ID NO: 179

           The protein of SEQ ID NO: 179 encoded by the extended cDNA SEQ ID NO: 78 found in testis exhibits

15   homology to part of mammalian colipase precursors. Colipases are secreted cofactors for pancreatic lipases that allow

the lipase to anchor at the water-lipid interface. Colipase plays a crucial role in the intestinal digestion and absorption of

dietary fats. The 5 cysteines characteristic for this protein family are conserved in the protein of the invention although

the colipase PROSITE signature is not.

           Taken together, these data suggest that the protein of SEQ ED NO: 179 may play a role in the lipid metabolism

20   and/or in male fertility. Thus, this protein may be useful in diagnosing and/or treating several types of disorders

including, but not limited to, hyperlipidemia, hypercholesterolemia, atherosclerosis, cardiovascular disorders such as

coronary heart disease, and neurodegenerative disorders such as Alzheimer's disease or dementia, and disorders linked to

male fertility.

25   ## Protein of SEQ ID NO: 227

           The protein of SEQ ID NO: 227 encoded by the extended cDNA SEQ ID NO: 126 exhibits extensive homology to

the ATP binding region of a whole family of serine/threonine protein kinases belonging to the CDC2/CDC28 subfamily.

The PROSITE signature characteristic for this domain is present in the protein of the invention from positions 10 to 34.

           Taken together, these data suggest that the protein of SEQ ED NO: 158 may bind ATP, and even be a protein

30   kinase. Thus, this protein may be useful in diagnosing and/or treating several types of disorders including, but not

limited to, cancer, neurodegenerative diseases, cardiovascular disorders, hypertension, renal injury and repair and septic

shock.

-97-

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims.

5   As discussed above, the extended cDNAs of the present invention or portions thereof can be used for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to

10   compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination for expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or

15   potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

        The proteins or polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit

20   another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other

25   protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

        Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

        Methods for performing the uses listed above are well known to those skilled in the art. References disclosing

30   such methods include without limitation "Molecular Cloning; A Laboratory Manual", 2d ed., Cole Spring Harbor Laboratory Press, Sambrook, J., E.F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology; Guide to Molecular Cloning Techniques", Academic Press, Berger, S.L. and A.R. Kimmel eds., 1987.

        Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a

-98-

nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

-99-

SEQUENCE LISTING FREE TEXT

The following free text appears in the accompanying Sequence Listing:

In vitro transcription product

oligonucleotide

5   promoter

transcription start site

Von Heijne matrix

Score

matinspector prediction

10 name

TABLE I

SEQ ID NO. in Present application	Provisional Application Disclosing Sequence	SEQ ID NO. in provisional application
40	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	51
41	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	72
42	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	52
43	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	78
44	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	73
45	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	41
46	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	67
47	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	82
48	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	80
49	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	81
50	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	53
51	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	54
52	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	195
53	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	44
54	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	46
55	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	68
56	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	48
57	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	55
58	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	49
59	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	50
60	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	97
61	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	51
62	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	69
63	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	49
64	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	199
65	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	53
66	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	57
67	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	54
68	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	55
69	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	58
70	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	59

CONT. TABLE I

71	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	60
72	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	112
73	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	52
74	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	59
75	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	60
76	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	136
77	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	75
78	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	61
79	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	61
80	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	130
81	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	65
82	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	54
83	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	78
84	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	63
85	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	65
86	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	152
87	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	66
88	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	67
89	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	60
90	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	68
91	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	61
92	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	62
93	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	166
94	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	70
95	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	73
96	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	63
97	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	52
98	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	62
99	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	176
100	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	63
101	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	187
102	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	190
103	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	83
104	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	180
105	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	64
106	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	69

CONT. TABLE I

107	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	40
108	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	77
109	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	43
110	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	82
111	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	76
112	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	43
113	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	46
114	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	47
115	U.S. Provisional Patent Application Serial No. 60/066,677, filed Nov. 13, 1997	53
116	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	58
117	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	74
118	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	71
119	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	145
120	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	67
121	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	58
122	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	72
123	U.S. Provisional Patent Application Serial No. 60/074,121, filed Feb. 9, 1998	73
124	U.S. Provisional Patent Application Serial No. 60/081,563, filed Apr. 13, 1998	70
125	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	40
126	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	44
127	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	45
128	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	47
129	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	48
130	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	51
131	U.S. Provisional Patent Application Serial No. 60/066,677, filed Nov. 13, 1997	50
132	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	56
133	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	57
134	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	71
135	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	72
136	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	64
137	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	65
138	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	66
139	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	74
140	U.S. Provisional Patent Application Serial No. 60/096,116, filed Aug. 10, 1998	67
242	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	75
243	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	76

CONT. TABLE I

244	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	77
245	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	78
246	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	79
247	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	80
248	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	81
249	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	82
250	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	83
251	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	84
252	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	85
253	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	86
254	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	87
255	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	88
256	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	89
257	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	90
258	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	91
259	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	92
260	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	93
261	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	94
262	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	95
263	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	96
264	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	97
265	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	98
266	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	99
267	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	100
268	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	101
269	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	102
270	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	103
271	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	104
272	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	105
273	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	106
274	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	107
275	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	108
276	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	109
277	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	110
278	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	111
279	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	112

CONT. TABLE I

280	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	113
281	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	114
282	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	115
283	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	116
284	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	117
285	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	118
286	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	119
287	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	120
288	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	121
289	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	122
290	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	123
291	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	124
292	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	125
293	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	126
294	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	127
295	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	128
296	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	129
297	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	130
298	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	131
299	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	132
300	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	133
301	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	134
302	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	135
303	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	136
304	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	137
305	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	138
306	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	139
307	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	140
308	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	141
309	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	142
310	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	143
311	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	144
312	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	145
313	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	146
314	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	147
315	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	148

CONT. TABLE I

316	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	149
317	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	150
318	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	151
319	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	152
320	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	153
321	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	154
322	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	155
323	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	156
324	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	157
325	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	158
326	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	159
327	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	160
328	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	161
329	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	162
330	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	163
331	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	164
332	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	165
333	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	166
334	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	167
335	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	168
336	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	169
337	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	170
338	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	171
339	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	172
340	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	173
341	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	174
342	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	175
343	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	176
344	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	177
345	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	178
346	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	179
347	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	180
348	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	181
349	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	182
350	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	183
351	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	184

CONT. TABLE I

352	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	185
353	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	186
354	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	187
355	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	188
356	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	189
357	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	190
358	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	191
359	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	192
360	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	193
361	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	194
362	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	195
363	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	196
364	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	197
365	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	1998
366	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	199
367	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	200
368	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	201
369	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	202
370	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	203
371	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	204
372	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	205
373	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	206
374	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	207
375	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	208
376	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	209
377	U.S. Provisional Patent Application Serial No. 60/069,957, filed Dec. 17, 1997	210

-107-

TABLE II : Parameters used for each step of EST analysis

Step	Program	Strand	Parameters	Identity (%))	Length (bp)
	Search Characteristics			**Selection Characteristics**	
Miscellaneous	Blastn	both	S=61 X=16	90	17
tRNA	Fasta	both	.	80	60
rRNA	Blastn	both	S=108	80	40
mtRNA	Blastn	both	S=108	80	40
Procaryotic	Blastn	both	S=144	90	40
Fungal	Blastn	both	S=144	90	40
Alu	fasta*	both	.	70	40
L1	Blastn	both	S=72	70	40
Repeats	Blastn	both	S=72	70	40
Promoters	Blastn	top	S=54 X=16	90	15⊥
Vertebrate	fasta*	both	S=108	90	30
ESTs	Blatsn	both	S=108 X=16	90	30
Proteins	blastxη	top	E=0.001	.	.

* use "Quick Fast" Database Scanner
⊥ alignment further constrained to begin closer than 10bp to EST\5' end
5 η using BLOSUM62 substitution matrix

TABLE III: Parameters used for each step of extended cDNA analysis

Step	Search characteristics		Selection characteristics			
	Program	Strand	Parameters	Identity (%)	Length (bp)	Comments
miscellaneous [*]	FASTA	both	.	90	15	
tRNA [¹]	FASTA	both	.	80	90	
rRNA [¹]	BLASTN	both	S=108	80	40	
mtRNA [¹]	BLASTN	both	S=108	80	40	
Procaryotic [¹]	BLASTN	both	S=144	90	40	
Fungal [*]	BLASTN	both	S=144	90	40	
Alu [*]	BLASTN	both	S=72	70	40	max 5 matches, masking
L1 [¹]	BLASTN	both	S=72	70	40	max 5 matches, masking
Repeats [¹]	BLASTN	both	S=72	70	40	masking
PolyA	BLAST2N	top	W=6,S=10,E=1000	90	8	in the last 20 nucleotides
Polyadenylati on signal	.	top	AATAAA allowing 1 mismatch			in the 50 nucleotides preceding the 5' end of the polA
Vertebrate [*]	BLASTN then FASTA	both	.	90 then 70	30	first BLASTN and then FASTA on matching sequences
ESTs [*]	BLAST2N	both	.	90	30	
Geneseq	BLASTN	both	W=8, B=10	90	30	
ORF	BLASTP	top	W=8, B=10	.	.	on ORF proteins, max 10 matches
Proteins [*]	BLASTX	top	E=0.001	70	30	

[¹] steps common to EST analysis and using the same algorithms and parameters
5 [*] steps also used in EST analysis but with different algorithms and/or parameters

TABLE IV

Id	FCS Location	SigPep Location	Mature Polypeptide Location	Stop Codon Location	PolyA Signal Location	PolyA Site Location
40	7 through 471	7 through 99	100 through 471	472	537 through 542	554 through 568
41	168 through 332	·	168 through 332	333	557 through 562	·
42	51 through 251	51 through 110	111 through 251	252	849 through 854	882 through 895
43	20 through 613	20 through 82	83 through 613	614	·	·
44	12 through 416	12 through 86	87 through 416	417	425 through 430	445 through 458
45	276 through 1040	276 through 485	486 through 1040	1041	·	2024 through 2036
46	443 through 619	443 through 589	590 through 619	620	·	1267 through 1276
47	206 through 747	·	206 through 747	·	·	·
48	36 through 521	36 through 104	105 through 521	522	528 through 533	548 through 561
49	36 through 395	36 through 104	105 through 395	396	599 through 604	619 through 632
50	21 through 41	·	21 through 41	42	328 through 333	357 through 370
51	35 through 631	35 through 160	161 through 631	632	901 through 906	979 through 994
52	271 through 399	·	271 through 399	400	·	·
53	103 through 252	103 through 213	214 through 252	253	·	588 through 597
54	2 through 460	·	2 through 460	461	713 through 718	735 through 748
55	31 through 231	·	31 through 231	232	769 through 774	690 through 703
56	305 through 565	·	305 through 565	566	694 through 699	713 through 725
57	124 through 873	124 through 378	379 through 873	874	1673 through 1678	1694 through 1705
58	135 through 206	·	135 through 206	207	850 through 855 ·	1056 through 1069
59	135 through 818	·	135 through 818	819	909 through 914	1071 through 1084
60	33 through 290	33 through 92	93 through 290	291	·	·
61	485 through 616	·	485 through 616	617	·	669 through 682
62	54 through 995	54 through 227	228 through 995	996	1130 through 1135	1181 through 1191
63	657 through 923	657 through 896	897 through 923	924	957 through 962	974 through 1008
64	18 through 311	18 through 62	63 through 311	312	·	·
65	151 through 426	151 through 258	259 through 426	427	505 through 510	527 through 538
66	10 through 1062	10 through 57	58 through 1062	1063	1710 through 1715	1735 through 1747
67	78 through 491	78 through 218	219 through 491	492	1652 through 1657	1673 through 1686
68	69 through 371	69 through 287	288 through 371	372	510 through 515	530 through 542
69	2 through 757	2 through 205	206 through 757	758	·	1160 through 1174
70	2 through 1051	2 through 205	206 through 1051	1052	1248 through 1253	1272 through 1285
71	2 through 1171	2 through 205	206 through 1171	1172	1368 through 1373	1386 through 1398
72	42 through 611	42 through 287	288 through 611	612	787 through 792	808 through 821
73	62 through 916	62 through 757	758 through 916	·	·	904 through 916
74	62 through 520	·	62 through 520	521	1124 through 1129	1141 through 1153
75	21 through 167	·	21 through 167	168	·	·
76	22 through 318	22 through 93	94 through 318	319	497 through 502	516 through 526
77	8 through 292	8 through 118	119 through 292	293	317 through 322	339 through 352
78	16 through 378	16 through 84	85 through 378	379	502 through 507	522 through 542

CONT. TABLE IV

79	57 through 233	·	57 through 233	·	·	·
80	83 through 340	83 through 124	125 through 340	341	573 through 578	607 through 660
81	47 through 541	47 through 220	221 through 541	542	·	597 through 605
82	46 through 285	46 through 150	151 through 285	286	364 through 369	385 through 396
83	22 through 240	22 through 84	85 through 240	241	397 through 402	421 through 432
84	89 through 382	·	89 through 382	383	·	408 through 420
85	80 through 415	80 through 142	143 through 415	416	471 through 476	488 through 501
86	152 through 361	152 through 283	284 through 361	362	·	·
87	32 through 307	32 through 70	71 through 307	308	1240 through 1245	1261 through 1272
88	114 through 734	114 through 239	240 through 734	735	768 through 773	793 through 804
89	199 through 802	·	199 through 802	·	780 through 785	791 through 802
90	38 through 1174	38 through 148	149 through 1174	1175	1452 through 1457	1478 through 1490
91	26 through 361	·	26 through 361	·	·	350 through 361
92	3 through 131	·	3 through 131	132	·	591 through 605
93	33 through 185	33 through 80	81 through 185	186	570 through 575	586 through 591
94	184 through 915	184 through 237	238 through 915	916	1119 through 1124	1139 through 1150
95	58 through 1116	58 through 159	160 through 1116	1117	1486 through 1491	1504 through 1513
96	327 through 417	·	327 through 417	·	·	404 through 417
97	63 through 398	63 through 206	207 through 398	399	·	·
98	2 through 163	·	2 through 163	164	488 through 493	511 through 522
99	13 through 465	13 through 75	76 through 465	466	·	·
100	20 through 703	20 through 94	95 through 703	704	1000 through 1005	1023 through 1041
101	103 through 294	103 through 243	244 through 294	295	·	·
102	81 through 518	81 through 173	174 through 518	519	·	·
103	66 through 326	·	66 through 326	327	1066 through 1071	1087 through 1098
104	170 through 289	170 through 250	251 through 289	290	·	·
105	36 through 497	·	36 through 497	498	650 through 655	663 through 685
106	18 through 320	·	18 through 320	321	539 through 544	542 through 554
107	71 through 1438	71 through 136	137 through 1438	1439	1644 through 1649	1665 through 1678
108	25 through 318	25 through 75	76 through 318	319	452 through 457	482 through 494
109	84 through 332	84 through 170	171 through 332	333	·	702 through 714
110	32 through 718	32 through 100	101 through 718	719	770 through 775	793 through 805
111	26 through 481	26 through 88	89 through 481	482	755 through 760	775 through 787
112	26 through 562	26 through 187	188 through 562	563	·	·
113	4 through 810	4 through 279	280 through 810	811	858 through 863	881 through 893
114	55 through 459	55 through 120	121 through 459	460	1444 through 1449	1462 through 1475
115	48 through 248	48 through 161	162 through 248	249	283 through 288	308 through 321
116	25 through 399	25 through 186	187 through 399	400	·	·
117	10 through 1137	10 through 72	73 through 1137	1138	1144 through 1149	1162 through 1173
118	72 through 704	72 through 161	162 through 704	705	772 through 777	·
119	44 through 505	44 through 223	224 through 505	506	·	·
120	25 through 393	25 through 150	151 through 393	394	734 through 739	757 through 770

CONT. TABLE IV

121	58 through 1095	58 through 114	115 through 1095	1096	.	1202 through 1213
122	31 through 660	31 through 90	91 through 660	661	1288 through 1293	1307 through 1318
123	31 through 582	31 through 90	91 through 582	583	816 through 821	840 through 853
124	15 through 695	15 through 80	81 through 695	696	795 through 800	814 through 826
125	74 through 295	74 through 196	197 through 295	296	545 through 550	561 through 571
126	440 through 659	.	440 through 659	.	601 through 606	.
127	38 through 283	38 through 85	86 through 283	284	257 through 262	.
128	121 through 477	121 through 288	289 through 477	.	.	.
129	2 through 163	.	2 through 163	164	292 through 297	310 through 323
130	46 through 675	46 through 87	88 through 675	676	1364 through 1369	1383 through 1392
131	62 through 385	.	62 through 385	386	974 through 979	987 through 999
132	422 through 550	422 through 475	476 through 550	551	.	714 through 725
133	124 through 231	.	124 through 231	232	.	387 through 400
134	131 through 1053	131 through 169	170 through 1053	.	1019 through 1024	.
135	86 through 403	86 through 181	182 through 403	404	1097 through 1102	1117 through 1128
136	37 through 162	37 through 93	94 through 162	163	224 through 229	243 through 254
137	31 through 381	31 through 90	91 through 381	382	.	875 through 886
138	46 through 579	46 through 156	157 through 579	580	.	.
139	92 through 471	92 through 172	173 through 471	.	454 through 459	458 through 471
140	154 through 675	154 through 498	499 through 675	676	819 through 824	838 through 849
242	18 through 173	18 through 77	78 through 173	174	864 through 869	882 through 893
243	17 through 595	17 through 85	86 through 595	596	820 through 825	840 through 851
244	89 through 334	89 through 130	131 through 334	335	462 through 467	484 through 495
245	21 through 614	21 through 83	84 through 614	615	849 through 854	873 through 884
246	94 through 573	94 through 258	259 through 573	574	862 through 867	886 through 897
247	74 through 397	74 through 127	128 through 397	398	472 through 477	507 through 518
248	51 through 242	51 through 116	117 through 242	243	319 through 324	339 through 350
249	111 through 191	111 through 155	156 through 191	192	965 through 970	986 through 996
250	45 through 602	45 through 107	108 through 602	603	828 through 833	850 through 860
251	24 through 560	24 through 101	102 through 560	561	563 through 568	583 through 593
252	109 through 558	109 through 273	274 through 558	559	.	1104 through 1114
253	128 through 835	128 through 220	221 through 835	836	1145 through 1150	1170 through 1181
254	59 through 505	59 through 358	359 through 505	506	1042 through 1047	1062 through 1073
255	1 through 207	1 through 147	148 through 207	208	784 through 789	807 through 818
256	12 through 734	12 through 101	102 through 734	735	914 through 919	961 through 971
257	378 through 518	378 through 467	468 through 518	519	607 through 612	628 through 640
258	110 through 304	110 through 193	194 through 304	305	708 through 713	732 through 743
259	201 through 419	201 through 272	273 through 419	420	601 through 606	627 through 637
260	123 through 302	123 through 176	177 through 302	303	1279 through 1284	1301 through 1312
261	98 through 673	98 through 376	377 through 673	674	.	1025 through 1035
262	17 through 463	17 through 232	233 through 463	464	657 through 662	684 through 696
263	263 through 481	263 through 322	323 through 481	482	.	858 through 868

CONT. TABLE IV

264	42 through 299	42 through 101	102 through 299	300	.	762 through 775
265	198 through 431	198 through 260	261 through 431	432	.	1064 through 1074
266	279 through 473	279 through 362	363 through 473	474	944 through 949	970 through 981
267	12 through 644	12 through 92	93 through 644	645	1002 through 1007	1020 through 1031
268	91 through 459	91 through 330	331 through 459	460	.	1271 through 1281
269	70 through 327	70 through 147	148 through 327	328	1741 through 1746	1763 through 1774
270	12 through 497	12 through 104	105 through 497	498	935 through 940	955 through 967
271	90 through 383	90 through 200	201 through 383	384	609 through 614	632 through 643
272	332 through 541	332 through 376	377 through 541	542	739 through 744	761 through 773
273	43 through 222	43 through 177	178 through 222	223	530 through 535	555 through 566
274	115 through 231	115 through 180	181 through 231	232	419 through 424	445 through 455
275	232 through 384	232 through 300	301 through 384	385	650 through 655	662 through 673
276	143 through 427	143 through 286	287 through 427	428	606 through 611	628 through 639
277	284 through 463	294 through 379	380 through 463	464	.	762 through 772
278	162 through 671	162 through 398	399 through 671	672	805 through 810	830 through 840
279	63 through 632	63 through 308	309 through 632	633	808 through 813	829 through 840
280	21 through 362	21 through 200	201 through 362	363	821 through 826	838 through 849
281	21 through 503	21 through 344	345 through 503	504	1305 through 1310	1330 through 1341
282	1 through 201	1 through 63	64 through 201	202	637 through 642	660 through 671
283	39 through 1034	39 through 134	135 through 1034	1035	1566 through 1571	1587 through 1597
284	69 through 263	69 through 125	126 through 263	264	1173 through 1178	1196 through 1205
285	115 through 285	115 through 204	205 through 285	286	505 through 510	525 through 536
286	90 through 344	90 through 140	141 through 344	345	500 through 505	515 through 527
287	57 through 311	57 through 107	108 through 311	312	467 through 472	482 through 493
288	96 through 302	96 through 182	183 through 302	303	.	501 through 514
289	161 through 526	161 through 328	329 through 526	527	.	799 through 811
290	210 through 332	210 through 299	300 through 332	333	594 through 599	613 through 625
291	212 through 361	212 through 319	320 through 361	362	650 through 655	673 through 684
292	75 through 482	75 through 128	129 through 482	483	595 through 600	618 through 627
293	50 through 631	50 through 244	245 through 631	632	777 through 782	801 through 812
294	154 through 576	154 through 360	361 through 576	577	737 through 742	763 through 775
295	154 through 897	154 through 360	361 through 897	898	1017 through 1022	1044 through 1054
296	146 through 292	146 through 253	254 through 292	293	395 through 400	433 through 444
297	126 through 383	126 through 167	168 through 383	384	726 through 731	743 through 754
298	66 through 497	66 through 239	240 through 497	498	594 through 599	618 through 629
299	49 through 411	49 through 96	97 through 411	412	732 through 737	750 through 763
300	49 through 534	49 through 96	97 through 534	535	593 through 598	612 through 623
301	86 through 415	86 through 145	146 through 415	416	540 through 545	560 through 571
302	56 through 268	56 through 100	101 through 268	269	584 through 589	601 through 612
303	32 through 328	32 through 103	104 through 328	329	508 through 513	528 through 539
304	21 through 527	21 through 95	96 through 527	528	921 through 926	953 through 963
305	147 through 647	147 through 374	375 through 647	648	.	668 through 681

CONT. TABLE IV

306	262 through 471	262 through 306	307 through 471	472	663 through 668	682 through 693
307	74 through 1216	74 through 172	173 through 1216	1217	1627 through 1632	1640 through 1652
308	48 through 164	48 through 89	90 through 164	165	482 through 487	505 through 517
309	185 through 334	185 through 295	296 through 334	335	355 through 360	392 through 405
310	195 through 347	195 through 272	273 through 347	348	1037 through 1042	1071 through 1082
311	90 through 815	90 through 179	180 through 815	816	883 through 888	905 through 916
312	52 through 513	52 through 231	232 through 513	514	553 through 558	572 through 583
313	172 through 438	172 through 354	355 through 438	439	682 through 687	685 through 697
314	148 through 366	148 through 225	226 through 366	367	770 through 775	792 through 803
315	175 through 336	175 through 276	277 through 336	337	.	812 through 823
316	191 through 553	191 through 304	305 through 553	554	766 through 771	804 through 817
317	106 through 603	106 through 216	217 through 603	604	.	1102 through 1112
318	47 through 586	47 through 124	125 through 586	587	1583 through 1588	1614 through 1623
319	99 through 371	99 through 290	291 through 371	372	491 through 496	513 through 524
320	44 through 814	44 through 112	113 through 814	815	.	978 through 989
321	3 through 581	3 through 182	183 through 581	582	.	1006 through 1016
322	107 through 427	107 through 190	191 through 427	428	499 through 504	516 through 529
323	45 through 407	45 through 83	84 through 407	408	1008 through 1013	1032 through 1042
324	201 through 332	201 through 251	252 through 332	333	.	869 through 880
325	217 through 543	217 through 255	256 through 543	544	.	1206 through 1217
326	18 through 446	18 through 140	141 through 446	447	930 through 935	948 through 959
327	29 through 724	29 through 118	119 through 724	725	886 through 891	910 through 920
328	404 through 586	404 through 466	467 through 586	587	1304 through 1309	1334 through 1344
329	331 through 432	331 through 387	388 through 432	433	548 through 553	573 through 585
330	59 through 703	59 through 220	221 through 703	704	886 through 891	903 through 914
331	672 through 752	672 through 722	723 through 752	753	.	1150 through 1161
332	57 through 311	57 through 128	129 through 311	312	332 through 337	351 through 363
333	80 through 232	80 through 127	128 through 232	233	617 through 622	634 through 645
334	91 through 291	91 through 219	220 through 291	292	367 through 372	389 through 400
335	196 through 384	196 through 240	241 through 384	385	461 through 466	485 through 496
336	54 through 590	54 through 227	228 through 590	591	.	955 through 965
337	133 through 846	133 through 345	346 through 846	847	.	890 through 901
338	138 through 671	138 through 248	249 through 671	672	1319 through 1324	1338 through 1347
339	124 through 411	124 through 186	187 through 411	412	948 through 953	971 through 983
340	372 through 494	372 through 443	444 through 494	495	708 through 713	732 through 745
341	112 through 450	112 through 192	193 through 450	451	1053 through 1058	1095 through 1106
342	117 through 866	117 through 170	171 through 866	367	1159 through 1164	1178 through 1190
343	13 through 465	13 through 75	76 through 465	466	1035 through 1040	1060 through 1070
344	2 through 718	2 through 76	77 through 718	719	1170 through 1175	1203 through 1213
345	86 through 709	86 through 361	362 through 709	710	943 through 948	963 through 973
346	63 through 320	63 through 179	180 through 320	321	771 through 776	799 through 810
347	299 through 418	299 through 379	380 through 418	419	739 through 744	762 through 771

CONT. TABLE IV

348	186 through 380	186 through 233	234 through 380	381	383 through 388	396 through 409
349	69 through 458	69 through 233	234 through 458	459	564 through 569	602 through 613
350	12 through 638	12 through 263	264 through 638	639	951 through 956	975 through 985
351	282 through 389	282 through 332	333 through 389	390	1413 through 1418	1437 through 1447
352	208 through 339	208 through 294	295 through 339	340	.	1631 through 1641
353	69 through 557	69 through 224	225 through 557	558	849 through 854	870 through 883
354	134 through 325	134 through 274	275 through 325	326	.	718 through 728
355	78 through 731	78 through 227	228 through 731	732	.	1002 through 1013
356	46 through 693	46 through 90	91 through 693	694	937 through 942	962 through 973
357	126 through 527	126 through 182	183 through 527	528	834 through 839	856 through 867
358	66 through 320	66 through 113	114 through 320	321	490 through 495	508 through 519
359	73 through 948	73 through 159	160 through 948	949	.	1016 through 1028
360	69 through 434	69 through 236	237 through 434	435	419 through 424	441 through 452
361	628 through 804	628 through 711	712 through 804	805	.	864 through 875
362	70 through 366	70 through 108	109 through 366	367	496 through 501	521 through 531
363	70 through 366	70 through 108	109 through 366	367	.	1233 through 1244
364	111 through 434	111 through 185	.186 through 434	435	.	618 through 631
365	19 through 567	19 through 63	64 through 567	568	749 through 754	771 through 781
366	19 through 312	19 through 63	64 through 312	313	896 through 901	921 through 931
367	64 through 612	64 through 234	235 through 612	613	.	839 through 849
368	39 through 458	39 through 80	81 through 458	459	613 through 618	633 through 644
369	9 through 185	9 through 50	51 through 185	186	.	906 through 918
370	14 through 316	14 through 121	122 through 316	317	442 through 447	458 through 471
371	70 through 1092	70 through 234	235 through 1092	1093	1475 through 1480	1493 through 1504
372	274 through 597	274 through 399	400 through 597	598	731 through 736	754 through 765
373	230 through 469	230 through 307	308 through 469	470	1004 through 1009	1027 through 1040
374	72 through 545	72 through 203	204 through 545	546	.	1151 through 1162
375	36 through 425	36 through 119	120 through 425	426	1215 through 1220	1240 through 1250
376	155 through 751	155 through 340	341 through 751	752	912 through 917	937 through 947
377	46 through 585	46 through 120	121 through 585	586	584 through 589	606 through 619

TABLE V

Id	Full Length Polypeptide Location	Signal Peptide Location	Mature Polypeptide Location
141	-31 through 124	-31 through -1	1 through 124
142	1 through 55	.	1 through 55
143	-20 through 47	-20 through -1	1 through 47
144	-21 through 177	-21 through -1	1 through 177
145	-25 through 110	-25 through -1	1 through 110
146	-70 through 185	-70 through -1	1 through 185
147	-49 through 10	-49 through -1	1 through 10
148	1 through 180	.	1 through 180
149	-23 through 139	-23 through -1	1 through 139
150	-23 through 97	-23 through -1	1 through 97
151	1 through 7	.	1 through 7
152	-42 through 157	-42 through -1	1 through 157
153	1 through 43	.	1 through 43
154	-37 through 13	-37 through -1	1 through 13
155	1 through 153	.	1 through 153
156	1 through 67	.	1 through 67
157	1 through 87	.	1 through 87
158	-85 through 165	-85 through -1	1 through 165
159	1 through 24	.	1 through 24
160	1 through 228	.	1 through 228
161	-20 through 66	-20 through -1	1 through 66
162	1 through 44	.	1 through 44
163	-58 through 256	-58 through -1	1 through 256
164	-80 through 9	-80 through -1	1 through 9
165	-15 through 83	-15 through -1	1 through 83
166	-36 through 56	-36 through -1	1 through 56
167	-16 through 335	-16 through -1	1 through 335
168	-47 through 91	-47 through -1	1 through 91
169	-73 through 28	-73 through -1	1 through 28
170	-68 through 184	-68 through -1	1 through 184
171	-68 through 282	-68 through -1	1 through 282
172	-68 through 322	-68 through -1	1 through 322
173	-82 through 108	-82 through -1	1 through 108
174	-232 through 53	-232 through -1	1 through 53
175	1 through 153	.	1 through 153
176	1 through 49	.	1 through 49
177	-24 through 75	-24 through -1	1 through 75
178	-37 through 58	-37 through -1	1 through 58
179	-23 through 98	-23 through -1	1 through 98
180	1 through 59	.	1 through 59
181	-14 through 72	-14 through -1	1 through 72
182	-58 through 107	-58 through -1	1 through 107
183	-35 through 45	-35 through -1	1 through 45
184	-21 through 52	-21 through -1	1 through 52
185	1 through 98	.	1 through 98
186	-21 through 91	-21 through -1	1 through 91
187	-44 through 26	-44 through -1	1 through 26
188	-13 through 79	-13 through -1	1 through 79
189	-42 through 165	-42 through -1	1 through 165
190	1 through 201	.	1 through 201

CONT. TABLE V

191	-37 through 342	-37 through -1	1 through 342
192	1 through 112	.	1 through 112
193	1 through 43	.	1 through 43
194	-16 through 35	-16 through -1	1 through 35
195	-18 through 226	-18 through -1	1 through 226
196	-34 through 319	-34 through -1	1 through 319
197	1 through 30	.	1 through 30
198	-48 through 64	-48 through -1	1 through 64
199	1 through 54	.	1 through 54
200	-21 through 130	-21 through -1	1 through 130
201	-25 through 203	-25 through -1	1 through 203
202	-47 through 17	-47 through -1	1 through 17
203	-31 through 115	-31 through -1	1 through 115
204	1 through 87	.	1 through 87
205	-27 through 13	-27 through -1	1 through 13
206	1 through 154	.	1 through 154
207	1 through 101	.	1 through 101
208	-22 through 434	-22 through -1	1 through 434
209	-17 through 81	-17 through -1	1 through 81
210	-29 through 54	-29 through -1	1 through 54
211	-23 through 206	-23 through -1	1 through 206
212	-21 through 131	-21 through -1	1 through 131
213	-54 through 125	-54 through -1	1 through 125
214	-92 through 177	-92 through -1	1 through 177
215	-22 through 113	-22 through -1	1 through 113
216	-38 through 29	-38 through -1	1 through 29
217	-54 through 71	-54 through -1	1 through 71
218	-21 through 355	-21 through -1	1 through 355
219	-30 through 181	-30 through -1	1 through 181
220	-60 through 94	-60 through -1	1 through 94
221	-42 through 81	-42 through -1	1 through 81
222	-19 through 327	-19 through -1	1 through 327
223	-20 through 190	-20 through -1	1 through 190
224	-20 through 164	-20 through -1	1 through 164
225	-22 through 205	-22 through -1	1 through 205
226	-41 through 33	-41 through -1	1 through 33
227	1 through 73	.	1 through 73
228	-16 through 66	-16 through -1	1 through 66
229	-56 through 63	-56 through -1	1 through 63
230	1 through 54	.	1 through 54
231	-14 through 196	-14 through -1	1 through 196
232	1 through 108	.	1 through 108
233	-18 through 25	-18 through -1	1 through 25
234	1 through 36	.	1 through 36
235	-13 through 294	-13 through -1	1 through 294
236	-32 through 74	-32 through -1	1 through 74
237	-19 through 23	-19 through -1	1 through 23
238	-20 through 97	-20 through -1	1 through 97
239	-37 through 141	-37 through -1	1 through 141
240	-27 through 99	-27 through -1	1 through 99
241	-115 through 59	-115 through -1	1 through 59
378	-20 through 32	-20 through -1	1 through 32
379	-23 through 170	-23 through -1	1 through 170
380	-14 through 68	-14 through -1	1 through 68

CONT. TABLE V

381	-21 through 177	-21 through -1	1 through 177
382	-55 through 105	-55 through -1	1 through 105
383	-18 through 90	-18 through -1	1 through 90
384	-22 through 42	-22 through -1	1 through 42
385	-15 through 12	-15 through -1	1 through 12
386	-21 through 165	-21 through -1	1 through 165
387	-26 through 153	-26 through -1	1 through 153
388	-55 through 95	-55 through -1	1 through 95
389	-31 through 205	-31 through -1	1 through 205
390	-100 through 49	-100 through -1	1 through 49
391	-49 through 20	-49 through -1	1 through 20
392	-30 through 211	-30 through -1	1 through 211
393	-30 through 17	-30 through -1	1 through 17
394	-28 through 37	-28 through -1	1 through 37
395	-24 through 49	-24 through -1	1 through 49
396	-18 through 42	-18 through -1	1 through 42
397	-93 through 99	-93 through -1	1 through 99
398	-72 through 77	-72 through -1	1 through 77
399	-20 through 53	-20 through -1	1 through 53
400	-20 through 66	-20 through -1	1 through 66
401	-21 through 57	-21 through -1	1 through 57
402	-28 through 37	-28 through -1	1 through 37
403	-27 through 184	-27 through -1	1 through 184
404	-80 through 43	-80 through -1	1 through 43
405	-26 through 60	-26 through -1	1 through 60
406	-31 through 131	-31 through -1	1 through 131
407	-37 through 61	-37 through -1	1 through 61
408	-15 through 55	-15 through -1	1 through 55
409	-45 through 15	-45 through -1	1 through 15
410	-22 through 17	-22 through -1	1 through 17
411	-23 through 28	-23 through -1	1 through 28
412	-48 through 47	-48 through -1	1 through 47
413	-32 through 28	-32 through -1	1 through 28
414	-79 through 91	-79 through -1	1 through 91
415	-82 through 108	-82 through -1	1 through 108
416	-60 through 54	-60 through -1	1 through 54
417	-108 through 53	-108 through -1	1 through 53
418	-21 through 46	-21 through -1	1 through 46
419	-32 through 300	-32 through -1	1 through 300
420	-19 through 46	-19 through -1	1 through 46
422	-30 through 27	-30 through -1	1 through 27
423	-17 through 68	-17 through -1	1 through 68
424	-17 through 68	-17 through -1	1 through 68
425	-29 through 40	-29 through -1	1 through 40
426	-56 through 66	-56 through -1	1 through 66
427	-30 through 11	-30 through -1	1 through 11
428	-36 through 14	-36 through -1	1 through 14
429	-18 through 118	-18 through -1	1 through 118
430	-65 through 129	-65 through -1	1 through 129
431	-69 through 72	-69 through -1	1 through 72
432	-69 through 179	-69 through -1	1 through 179
433	-36 through 13	-36 through -1	1 through 13
434	-14 through 72	-14 through -1	1 through 72
435	-58 through 86	-58 through -1	1 through 86

CONT. TABLE V

436	-16 through 105	-16 through -1	1 through 105
437	-16 through 146	-16 through -1	1 through 146
438	-20 through 90	-20 through -1	1 through 90
439	-15 through 56	-15 through -1	1 through 56
440	-24 through 75	-24 through -1	1 through 75
441	-25 through 144	-25 through -1	1 through 144
442	-76 through 91	-76 through -1	1 through 91
443	-15 through 55	-15 through -1	1 through 55
444	-33 through 348	-33 through -1	1 through 348
445	-14 through 25	-14 through -1	1 through 25
446	-37 through 13	-37 through -1	1 through 13
447	-26 through 25	-26 through -1	1 through 25
448	-30 through 212	-30 through -1	1 through 212
449	-60 through 94	-60 through -1	1 through 94
450	-61 through 28	-61 through -1	1 through 28
451	-26 through 47	-26 through -1	1 through 47
452	-34 through 20	-34 through -1	1 through 20
453	-38 through 83	-38 through -1	1 through 83
454	-37 through 129	-37 through -1	1 through 129
455	-26 through 154	-26 through -1	1 through 154
456	-64 through 27	-64 through -1	1 through 27
457	-23 through 234	-23 through -1	1 through 234
458	-60 through 133	-60 through -1	1 through 133
459	-28 through 79	-28 through -1	1 through 79
460	-13 through 108	-13 through -1	1 through 108
461	-17 through 27	-17 through -1	1 through 27
462	-13 through 96	-13 through -1	1 through 96
463	-41 through 102	-41 through -1	1 through 102
464	-30 through 202	-30 through -1	1 through 202
465	-21 through 40	-21 through -1	1 through 40
466	-19 through 15	-19 through -1	1 through 15
467	-54 through 161	-54 through -1	1 through 161
468	-17 through 10	-17 through -1	1 through 10
469	-24 through 61	-24 through -1	1 through 61
470	-16 through 35	-16 through -1	1 through 35
471	-43 through 24	-43 through -1	1 through 24
472	-15 through 48	-15 through -1	1 through 48
473	-58 through 121	-58 through -1	1 through 121
474	-71 through 167	-71 through -1	1 through 167
475	-37 through 141	-37 through -1	1 through 141
476	-21 through 75	-21 through -1	1 through 75
477	-24 through 17	-24 through -1	1 through 17
478	-27 through 86	-27 through -1	1 through 86
479	-18 through 232	-18 through -1	1 through 232
480	-21 through 130	-21 through -1	1 through 130
481	-25 through 214	-25 through -1	1 through 214
482	-92 through 116	-92 through -1	1 through 116
483	-39 through 47	-39 through -1	1 through 47
484	-27 through 13	-27 through -1	1 through 13
485	-16 through 49	-16 through -1	1 through 49
486	-55 through 75	-55 through -1	1 through 75
487	-84 through 125	-84 through -1	1 through 125
488	-17 through 19	-17 through -1	1 through 19
489	-29 through 15	-29 through -1	1 through 15

490	·52 through 111	·52 through ·1	1 through 111
491	·47 through 17	·47 through ·1	1 through 17
492	·50 through 168	·50 through ·1	1 through 168
493	·15 through 201	·15 through ·1	1 through 201
494	·19 through 115	·19 through ·1	1 through 115
495	·16 through 69	·16 through ·1	1 through 69
496	·29 through 263	·29 through ·1	1 through 263
497	·56 through 66	·56 through ·1	1 through 66
498	·28 through 31	·28 through ·1	1 through 31
499	·13 through 86	·13 through ·1	1 through 86
500	·13 through 86	·13 through ·1	1 through 86
501	·25 through 83	·25 through ·1	1 through 83
502	·15 through 168	·15 through ·1	1 through 168
503	·15 through 83	·15 through ·1	1 through 83
504	·57 through 126	·57 through ·1	1 through 126
505	·14 through 126	·14 through ·1	1 through 126
506	·14 through 45	·14 through ·1	1 through 45
507	·36 through 65	·36 through ·1	1 through 65
508	·55 through 286	·55 through ·1	1 through 286
509	·42 through 66	·42 through ·1	1 through 66
510	·26 through 54	·26 through ·1	1 through 54
511	·44 through 114	·44 through ·1	1 through 114
512	·28 through 102	·28 through ·1	1 through 102
513	·62 through 137	·62 through ·1	1 through 137
514	·25 through 155	·25 through ·1	1 through 155

·120·

TABLE VI

Id	Collection refs	Deposit Name
40	ATCC # 98921	SignalTag 121-144
41	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
42	ATCC # 98921	SignalTag 121-144
43	ATCC # 98920	SignalTag 67-90
44	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
45	ATCC # 98920	SignalTag 67-90
46	ATCC # 98923	SignalTag 44-66
47	ATCC # 98920	SignalTag 67-90
48	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
49	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
50	ATCC # 98921	SignalTag 121-144
51	ATCC # 98921	SignalTag 121-144
52	ATCC # 98920	SignalTag 67-90
53	ATCC # 98923	SignalTag 44-66
54	ATCC # 98920	SignalTag 67-90
55	ATCC # 98920	SignalTag 67-90
56	ATCC # 98920	SignalTag 67-90
57	ATCC # 98921	SignalTag 121-144
58	ATCC # 98920	SignalTag 67-90
59	ATCC # 98920	SignalTag 67-90
60	ATCC # 98920	SignalTag 67-90
61	ATCC # 98923	SignalTag 44-66
62	ATCC # 98923	SignalTag 44-66
63	ATCC # 98923	SignalTag 44-66
64	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
65	ATCC # 98923	SignalTag 44-66
66	ATCC # 98921	SignalTag 121-144
67	ATCC # 98920	SignalTag 67-90
68	ATCC # 98920	SignalTag 67-90
69	ATCC # 98921	SignalTag 121-144
70	ATCC # 98921	SignalTag 121-144
71	ATCC # 98921	SignalTag 121-144
72	ATCC # 98922	SignalTag 91-94, 96, 37, 99-107, 109-112 et 114-120
73	ATCC # 98923	SignalTag 44-66

74	ATCC # 98923	SignalTag 44-66
75	ATCC # 98920	SignalTag 67-90
76	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
77	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
78	ATCC # 98921	SignalTag 121-144
79	ATCC # 98923	SignalTag 44-66
80	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
81	ATCC # 98921	SignalTag 121-144
82	ATCC # 98920	SignalTag 67-90
83	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
84	ATCC # 98923	SignalTag 44-66
85	ATCC # 98923	SignalTag 44-66
86	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
87	ATCC # 98923	SignalTag 44-66
88	ATCC # 98923	SignalTag 44-66
89	ATCC # 98923	SignalTag 44-66
90	ATCC # 98923	SignalTag 44-66
91	ATCC # 98923	SignalTag 44-66
92	ATCC # 98920	SignalTag 67-90
93	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
94	ATCC # 98923	SignalTag 44-66
95	ATCC # 98923	SignalTag 44-66
96	ATCC # 98920	SignalTag 67-90
97	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
98	ATCC # 98921	SignalTag 121-144
99	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
100	ATCC # 98921	SignalTag 121-144
101	ATCC # 98920	SignalTag 67-90
102	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
103	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
104	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
105	ATCC # 98921	SignalTag 121-144
106	ATCC # 98920	SignalTag 67-90
107	ATCC # 98920	SignalTag 67-90
108	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
109	ATCC # 98923	SignalTag 44-66
110	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120

-122-

111	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
112	ATCC # 98920	SignalTag 67-90
113	ATCC # 98920	SignalTag 67-90
114	ATCC # 98923	SignalTag 44-66
115	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
116	ATCC # 98920	SignalTag 67-90
117	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
118	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
119	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
120	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
121	ATCC # 98923	SignalTag 44-66
122	ATCC # 98920	SignalTag 67-90
123	ATCC # 98920	SignalTag 67-90
124	ATCC # 98922	SignalTag 91-94, 96, 97, 99-107, 109-112 et 114-120
125	ECACC # 98121506	SignalTag 11121998
126	ECACC # 98121506	SignalTag 11121998
127	ECACC # 98121506	SignalTag 11121998
128	ECACC # 98121506	SignalTag 11121998
129	ECACC # 98121506	SignalTag 11121998
130	ECACC # 98121506	SignalTag 11121998
131	ECACC # 98121506	SignalTag 11121998
132	ECACC # 98121506	SignalTag 11121998
133	ECACC # 98121506	SignalTag 11121998
134	ECACC # 98121506	SignalTag 11121998
135	ECACC # 98121506	SignalTag 11121998
136	ECACC # 98121506	SignalTag 11121998
137	ECACC # 98121506	SignalTag 11121998
138	ECACC # 98121506	SignalTag 11121998
139	ECACC # 98121506	SignalTag 11121998
140	ECACC # 98121506	SignalTag 11121998

·123·

TABLE VII

Internal designation number	SEQ ID NO	Type of sequence
20-5-2-C3-CL0_4	40	DNA
20-8-4-A11-CL2_6	41	DNA
21-1-4-F2-CL11_1	42	DNA
22-11-2-H9-CL1_1	43	DNA
25-7-3-D4-CL0_2	44	DNA
26-27-3-D7-CL0_1	45	DNA
26-35-4-H9-CL1_1	46	DNA
26-45-2-C4-CL2_6	47	DNA
27-1-2-B3-CL0_1	48	DNA
27-1-2-B3-CL0_2	49	DNA
27-19-3-G7-CL11_2	50	DNA
33-10-4-E2-CL13_4	51	DNA
33-10-4-H2-CL2_2	52	DNA
33-110-4-A5-CL1_1	53	DNA
33-13-1-C1-CL1_1	54	DNA
33-30-2-A6-CL0_1	55	DNA
33-35-4-F4-CL1_2	56	DNA
33-35-4-G1-CL1_2	57	DNA
33-36-3-E2-CL1_1	58	DNA
33-36-3-E2-CL1_2	59	DNA
33-36-3-F2-CL2_2	60	DNA
33-4-2-G5-CL2_1	61	DNA
33-49-1-H4-CL1_1	62	DNA
33-66-2-B10-CL4_1	63	DNA
33-97-4-G8-CL2_2	64	DNA
33-98-4-C1-CL1_3	65	DNA
47-14-1-C3-CL0_5	66	DNA
47-15-1-E11-CL0_1	67	DNA
47-15-1-H8-CL0_2	68	DNA
48-1-1-H7-CL0_1	69	DNA
48-1-1-H7-CL0_4	70	DNA
48-1-1-H7-CL0_5	71	DNA
48-3-1-H9-CL0_6	72	DNA
48-54-1-G9-CL2_1	73	DNA

48-54-1-G9-CL3_1	74	DNA
48-7-4-H2-CL2_2	75	DNA
51-11-3-D5-CL1_3	76	DNA
51-11-3-G9-CL0_1	77	DNA
51-15-4-A12-CL11_3	78	DNA
51-17-4-A4-CL3_1	79	DNA
51-2-3-F10-CL1_5	80	DNA
51-2-4-F5-CL11_2	81	DNA
51-27-4-F2-CL0_2	82	DNA
51-34-3-F8-CL0_2	83	DNA
57-1-4-E2-CL1_2	84	DNA
57-19-2-G8-CL2_1	85	DNA
57-27-3-G10-CL2_2	86	DNA
58-33-3-B4-CL1_2	87	DNA
58-34-3-C9-CL1_2	88	DNA
58-4-4-G2-CL2_1	89	DNA
58-48-1-G3-CL2_4	90	DNA
58-6-1-H4-CL1_1	91	DNA
60-12-1-E11-CL1_2	92	DNA
65-4-4-H3-CL1_1	93	DNA
74-5-1-E4-CL1_2	94	DNA
76-13-3-A9-CL1_2	95	DNA
76-16-1-D6-CL1_1	96	DNA
76-28-3-A12-CL1_5	97	DNA
76-42-2-F3-CL0_1	98	DNA
77-16-4-G3-CL1_3	99	DNA
77-39-4-H4-CL11_4	100	DNA
78-24-3-H4-CL2_1	101	DNA
78-27-3-D1-CL1_6	102	DNA
78-28-3-D2-CL0_2	103	DNA
78-7-1-G5-CL2_6	104	DNA
84-3-1-G10-CL11_6	105	DNA
58-48-4-E2-CL0_1	106	DNA
23-12-2-G6-CL1_2	107	DNA
25-8-4-B12-CL0_5	108	DNA
26-44-3-C5-CL2_1	109	DNA
27-1-2-B3-CL0_3	110	DNA

30·12·3·G5·CL0_1	111	DNA
33·106·2·F10·CL1_3	112	DNA
33·28·4·D1·CL0_1	113	DNA
33·31·3·C8·CL2_1	114	DNA
48·24·1·D2·CL3_2	115	DNA
48·46·4·A11·CL1_4	116	DNA
51·1·4·C1·CL0_2	117	DNA
51·39·3·H2·CL1_2	118	DNA
51·42·3·F9·CL1_1	119	DNA
51·5·3·G2·CL0_4	120	DNA
57·18·4·H5·CL2_1	121	DNA
76·23·3·G8·CL1_1	122	DNA
76·23·3·G8·CL1_3	123	DNA
78·8·3·E6·CL0_1	124	DNA
19·10·1·C2·CL1_3	125	DNA
33·11·1·B11·CL1_2	126	DNA
33·113·2·B8·CL1_2	127	DNA
33·19·1·C11·CL1_1	128	DNA
33·61·2·F6·CL0_2	129	DNA
47·4·4·C6·CL2_2	130	DNA
48·54·1·G9·CL1_1	131	DNA
51·43·3·G3·CL0_1	132	DNA
55·1·3·D11·CL0_1	133	DNA
58·14·2·D3·CL1_2	134	DNA
58·35·2·B6·CL2_3	135	DNA
76·18·1·F6·CL1_1	136	DNA
76·23·3·G8·CL2_2	137	DNA
76·30·3·B7·CL1_1	138	DNA
78·21·3·G7·CL2_1	139	DNA
58·45·4·B11·CL13_2	140	DNA
20·5·2·C3·CL0_4	141	PRT
20·8·4·A11·CL2_6	142	PRT
21·1·4·F2·CL11_1	143	PRT
22·11·2·H9·CL1_1	144	PRT
25·7·3·D4·CL0_2	145	PRT
26·27·3·D7·CL0_1	146	PRT
26·35·4·H9·CL1_1	147	PRT

26-45-2-C4-CL2_6	148	PRT
27-1-2-B3-CL0_1	149	PRT
27-1-2-B3-CL0_2	150	PRT
27-19-3-G7-CL11_2	151	PRT
33-10-4-E2-CL13_4	152	PRT
33-10-4-H2-CL2_2	153	PRT
33-110-4-A5-CL1_1	154	PRT
33-13-1-C1-CL1_1	155	PRT
33-30-2-A6-CL0_1	156	PRT
33-35-4-F4-CL1_2	157	PRT
33-35-4-G1-CL1_2	158	PRT
33-36-3-E2-CL1_1	159	PRT
33-36-3-E2-CL1_2	160	PRT
33-36-3-F2-CL2_2	161	PRT
33-4-2-G5-CL2_1	162	PRT
33-49-1-H4-CL1_1	163	PRT
33-66-2-B10-CL4_1	164	PRT
33-97-4-G8-CL2_2	165	PRT
33-98-4-C1-CL1_3	166	PRT
47-14-1-C3-CL0_5	167	PRT
47-15-1-E11-CL0_1	168	PRT
47-15-1-H8-CL0_2	169	PRT
48-1-1-H7-CL0_1	170	PRT
48-1-1-H7-CL0_4	171	PRT
48-1-1-H7-CL0_5	172	PRT
48-3-1-H9-CL0_6	173	PRT
48-54-1-G9-CL2_1	174	PRT
48-54-1-G9-CL3_1	175	PRT
48-7-4-H2-CL2_2	176	PRT
51-11-3-D5-CL1_3	177	PRT
51-11-3-G9-CL0_1	178	PRT
51-15-4-A12-CL11_3	179	PRT
51-17-4-A4-CL3_1	180	PRT
51-2-3-F10-CL1_5	181	PRT
51-2-4-F5-CL11_2	182	PRT
51-27-4-F2-CL0_2	183	PRT
51-34-3-F8-CL0_2	184	PRT

57-1-4-E2-CL1_2	185	PRT
57-19-2-G8-CL2_1	186	PRT
57-27-3-G10-CL2_2	187	PRT
58-33-3-B4-CL1_2	188	PRT
58-34-3-C9-CL1_2	189	PRT
58-4-4-G2-CL2_1	190	PRT
58-48-1-G3-CL2_4	191	PRT
58-6-1-H4-CL1_1	192	PRT
60-12-1-E11-CL1_2	193	PRT
65-4-4-H3-CL1_1	194	PRT
74-5-1-E4-CL1_2	195	PRT
76-13-3-A9-CL1_2	196	PRT
76-16-1-D6-CL1_1	197	PRT
76-28-3-A12-CL1_5	198	PRT
76-42-2-F3-CL0_1	199	PRT
77-16-4-G3-CL1_3	200	PRT
77-39-4-H4-CL11_4	201	PRT
78-24-3-H4-CL2_1	202	PRT
78-27-3-D1-CL1_6	203	PRT
78-28-3-D2-CL0_2	204	PRT
78-7-1-G5-CL2_6	205	PRT
84-3-1-G10-CL11_6	206	PRT
58-48-4-E2-CL0_1	207	PRT
23-12-2-G6-CL1_2	208	PRT
25-8-4-B12-CL0_5	209	PRT
26-44-3-C5-CL2_1	210	PRT
27-1-2-B3-CL0_3	211	PRT
30-12-3-G5-CL0_1	212	PRT
33-106-2-F10-CL1_3	213	PRT
33-28-4-D1-CL0_1	214	PRT
33-31-3-C8-CL2_1	215	PRT
48-24-1-D2-CL3_2	216	PRT
48-46-4-A11-CL1_4	217	PRT
51-1-4-C1-CL0_2	218	PRT
51-39-3-H2-CL1_2	219	PRT
51-42-3-F9-CL1_1	220	PRT
51-5-3-G2-CL0_4	221	PRT

57-18-4-H5-CL2_1	222	PRT
76-23-3-G8-CL1_1	223	PRT
76-23-3-G8-CL1_3	224	PRT
78-8-3-E6-CL0_1	225	PRT
19-10-1-C2-CL1_3	226	PRT
33-11-1-B11-CL1_2	227	PRT
33-113-2-B8-CL1_2	228	PRT
33-19-1-C11-CL1_1	229	PRT
33-61-2-F6-CL0_2	230	PRT
47-4-4-C6-CL2_2	231	PRT
48-54-1-G9-CL1_1	232	· PRT
51-43-3-G3-CL0_1	233	PRT
55-1-3-D11-CL0_1	234	PRT
58-14-2-D3-CL1_2	235	PRT
58-35-2-B6-CL2_3	236	PRT
76-18-1-F6-CL1_1	237	PRT
76-23-3-G8-CL2_2	238	PRT
76-30-3-B7-CL1_1	239	PRT
78-21-3-G7-CL2_1	240	PRT
58-45-4-B11-CL13_2	241	PRT
20-6-1-D11-FL2	242	DNA
20-8-4-A11-FL2	243	DNA
22-6-2-C1-FL2	244	DNA
22-11-2-H9-FL1	245	DNA
23-8-3-B1-FL1	246	DNA
24-3-3-C6-FL1	247	DNA
24-4-1-H3-FL1	248	DNA
26-45-2-C4-FL2	249	DNA
26-48-1-H10-FL1	250	DNA
26-49-1-A5-FL2	251	DNA
30-6-4-E3-FL3	252	DNA
33-6-1-G11-FL1	253	DNA
33-8-1-A3-FL2	254	DNA
33-11-3-C6-FL1	255	DNA
33-14-4-E1-FL1	256	DNA
33-21-2-D5-FL1	257	DNA
33-26-4-E10-FL1	258	DNA

·129·

33-27-1-E11-FL1	259	DNA
33-28-4-D1-FL1	260	DNA
33-28-4-E2-FL2	261	DNA
33-30-4-C4-FL1	262	DNA
33-35-4-F4-FL1	263	DNA
33-36-3-F2-FL2	264	DNA
33-52-4-F9-FL2	265	DNA
33-52-4-H3-FL1	266	DNA
33-59-1-B7-FL1	267	DNA
33-71-1-A8-FL1	268	DNA
33-72-2-B2-FL1	269	DNA
33-105-2-C3-FL1	270	DNA
33-107-4-C3-FL1	271	DNA
33-110-2-G4-FL1	272	DNA
47-7-4-D2-FL2	273	DNA
47-10-2-G12-FL1	274	DNA
47-14-3-D8-FL1	275	DNA
47-18-3-C2-FL1	276	DNA
47-18-3-G5-FL2	277	DNA
47-18-4-E3-FL2	278	DNA
48-3-1-H9-FL3	279	DNA
48-4-2-H3-FL1	280	DNA
48-6-1-C9-FL1	281	DNA
48-7-4-H2-FL2	282	DNA
48-8-1-D8-FL3	283	DNA
48-13-3-H8-FL1	284	DNA
48-19-3-A7-FL1	285	DNA
48-19-3-G1-FL1	286	DNA
48-25-4-D8-FL1	287	DNA
48-21-4-H4-FL1	288	DNA
48-26-3-B8-FL2	289	DNA
48-29-1-E2-FL1	290	DNA
48-31-3-F7-FL1	291	DNA
48-47-3-A5-FL1	292	DNA
51-1-1-G12-FL1	293	DNA
51-1-4-E9-FL3	294	DNA
51-1-4-E9-FL2	295	DNA

51·2·1·E10·FL1	296	DNA
51·2·3·F10·FL1	297	DNA
51·2·4·F5·FL1	298	DNA
51·3·3·B10·FL2	299	DNA
51·3·3·B10·FL3	300	DNA
51·7·3·G3·FL1	301	DNA
51·10·3·D11·FL1	302	DNA
51·11·3·D5·FL1	303	DNA
51·13·1·F7·FL3	304	DNA
51·15·4·H10·FL1	305	DNA
51·17·4·A4·FL1	306	DNA
51·18·1·C3·FL1	307	DNA
51·25·3·F3·FL1	308	DNA
51·27·1·E8·FL1	309	DNA
51·28·2·G1·FL2	310	DNA
51·39·3·H2·FL1	311	DNA
51·42·3·F9·FL1	312	DNA
51·44·4·H4·FL1	313	DNA
55·1·3·H10·FL1	314	DNA
55·5·4·A6·FL1	315	DNA
58·26·3·D1·FL1	316	DNA
57·18·1·D5·FL1	317	DNA
57·27·3·A11·FL1	318	DNA
57·27·3·G10·FL2	319	DNA
58·10·3·D12·FL1	320	DNA
58·11·1·G10·FL1	321	DNA
58·11·2·G8·FL2	322	DNA
58·36·3·A9·FL2	323	DNA
58·38·1·A2·FL2	324	DNA
58·38·1·E5·FL1	325	DNA
58·44·2·B3·FL3	326	DNA
58·45·3·H11·FL1	327	DNA
58·53·2·B12·FL2	328	DNA
59·9·4·A10·FL1	329	DNA
60·16·3·A6·FL1	330	DNA
60·17·3·G8·FL2	331	DNA
62·5·4·B10·FL1	332	DNA

65-4-4-H3-FL1	333	DNA
74-3-1-B9-FL1	334	DNA
76-4-1-G5-FL1	335	DNA
76-7-3-A12-FL1	336	DNA
76-16-4-C9-FL3	337	DNA
76-30-3-B7-FL1	338	DNA
77-5-1-C2-FL1	339	DNA
77-5-4-E7-FL1	340	DNA
77-11-1-A3-FL1	341	DNA
77-16-3-D7-FL1	342	DNA
77-16-4-G3-FL1	343	DNA
77-25-1-A6-FL1	344	DNA
77-26-2-F2-FL3	345	DNA
78-6-2-E3-FL2	346	DNA
78-7-1-G5-FL2	347	DNA
78-16-2-C2-FL1	348	DNA
78-18-3-B4-FL3	349	DNA
78-20-1-G11-FL1	350	DNA
78-22-3-E10-FL1	351	DNA
78-24-2-B8-FL1	352	DNA
78-24-3-A8-FL1	353	DNA
78-24-3-H4-FL2	354	DNA
78-25-1-F11-FL1	355	DNA
78-26-1-B5-FL1	356	DNA
78-27-3-D1-FL1	357	DNA
78-29-1-B2-FL1	358	DNA
78-29-4-B6-FL1	359	DNA
14-1-3-E6-FL1	360	DNA
30-9-1-G8-FL2	361	DNA
33-10-4-H2-FL2	362	DNA
33-10-4-H2-FL1	363	DNA
74-10-3-C9-FL2	364	DNA
33-97-4-G8-FL3	365	DNA
33-97-4-G8-FL2	366	DNA
33-104-4-H4-FL1	367	DNA
47-2-3-B3-FL1	368	DNA
47-37-4-G11-FL1	369	DNA

57-25-1-F10-FL2	370	DNA
58-19-3-D3-FL1	371	DNA
58-34-3-C9-FL2	372	DNA
58-48-4-E2-FL2	373	DNA
76-21-1-C4-FL1	374	DNA
78-26-2-H7-FL1	375	DNA
77-20-2-E11-FL1	376	DNA
47-1-3-F7-FL2	377	DNA
20-6-1-D11-FL2	378	PRT
20-8-4-A11-FL2	379	PRT
22-6-2-C1-FL2	380	PRT
22-11-2-H9-FL1	381	PRT
23-8-3-B1-FL1	382	PRT
24-3-3-C6-FL1	383	PRT
24-4-1-H3-FL1	384	PRT
26-45-2-C4-FL2	385	PRT
26-48-1-H10-FL1	386	PRT
26-49-1-A5-FL2	387	PRT
30-6-4-E3-FL3	388	PRT
33-6-1-G11-FL1	389	PRT
33-8-1-A3-FL2	390	PRT
33-11-3-C6-FL1	391	PRT
33-14-4-E1-FL1	392	PRT
33-21-2-D5-FL1	393	PRT
33-26-4-E10-FL1	394	PRT
33-27-1-E11-FL1	395	PRT
33-28-4-D1-FL1	396	PRT
33-28-4-E2-FL2	397	PRT
33-30-4-C4-FL1	398	PRT
33-35-4-F4-FL1	399	PRT
33-36-3-F2-FL2	400	PRT
33-52-4-F9-FL2	401	PRT
33-52-4-H3-FL1	402	PRT
33-59-1-B7-FL1	403	PRT
33-71-1-A8-FL1	404	PRT
33-72-2-B2-FL1	405	PRT
33-105-2-C3-FL1	406	PRT

33·107·4·C3·FL1	407	PRT
33·110·2·G4·FL1	408	PRT
47·7·4·D2·FL2	409	PRT
47·10·2·G12·FL1	410	PRT
47·14·3·D8·FL1	411	PRT
47·18·3·C2·FL1	412	PRT
47·18·3·G5·FL2	413	PRT
47·18·4·E3·FL2	414	PRT
48·3·1·H9·FL3	415	PRT
48·4·2·H3·FL1	416	PRT
48·6·1·C9·FL1	417	PRT
48·7·4·H2·FL2	418	PRT
48·8·1·D8·FL3	419	PRT
48·13·3·H8·FL1	420	PRT
48·19·3·A7·FL1	421	PRT
48·19·3·G1·FL1	422	PRT
48·25·4·D8·FL1	423	PRT
48·21·4·H4·FL1	424	PRT
48·26·3·B8·FL2	425	PRT
48·29·1·E2·FL1	426	PRT
48·31·3·F7·FL1	427	PRT
48·47·3·A5·FL1	428	PRT
51·1·1·G12·FL1	429	PRT
51·1·4·E9·FL3	430	PRT
51·1·4·E9·FL2	431	PRT
51·2·1·E10·FL1	432	PRT
51·2·3·F10·FL1	433	PRT
51·2·4·F5·FL1	434	PRT
51·3·3·B10·FL2	435	PRT
51·3·3·B10·FL3	436	PRT
51·7·3·G3·FL1	437	PRT
51·10·3·D11·FL1	438	PRT
51·11·3·D5·FL1	439	PRT
51·13·1·F7·FL3	440	PRT
51·15·4·H10·FL1	441	PRT
51·17·4·A4·FL1	442	PRT
51·18·1·C3·FL1	443	PRT

51-25-3-F3-FL1	444	PRT
51-27-1-E8-FL1	445	PRT
51-28-2-G1-FL2	446	PRT
51-39-3-H2-FL1	447	PRT
51-42-3-F9-FL1	448	PRT
51-44-4-H4-FL1	449	PRT
55-1-3-H10-FL1	450	PRT
55-5-4-A6-FL1	451	PRT
58-26-3-D1-FL1	452	PRT
57-18-1-D5-FL1	453	PRT
57-27-3-A11-FL1	454	PRT
57-27-3-G10-FL2	455	PRT
58-10-3-D12-FL1	456	PRT
58-11-1-G10-FL1	457	PRT
58-11-2-G8-FL2	458	PRT
58-36-3-A9-FL2	459	PRT
58-38-1-A2-FL2	460	PRT
58-38-1-E5-FL1	461	PRT
58-44-2-B3-FL3	462	PRT
58-45-3-H11-FL1	463	PRT
58-53-2-B12-FL2	464	PRT
59-9-4-A10-FL1	465	PRT
60-16-3-A6-FL1	466	PRT
60-17-3-G8-FL2	467	PRT
62-5-4-B10-FL1	468	PRT
65-4-4-H3-FL1	469	PRT
74-3-1-B9-FL1	470	PRT
76-4-1-G5-FL1	471	PRT
76-7-3-A12-FL1	472	PRT
76-16-4-C9-FL3	473	PRT
76-30-3-B7-FL1	474	PRT
77-5-1-C2-FL1	475	PRT
77-5-4-E7-FL1	476	PRT
77-11-1-A3-FL1	477	PRT
77-16-3-D7-FL1	478	PRT
77-16-4-G3-FL1	479	PRT
77-25-1-A6-FL1	480	PRT

77-26-2-F2-FL3	481	PRT
78-6-2-E3-FL2	482	PRT
78-7-1-G5-FL2	483	PRT
78-16-2-C2-FL1	484	PRT
78-18-3-B4-FL3	485	PRT
78-20-1-G11-FL1	486	PRT
78-22-3-E10-FL1	487	PRT
78-24-2-B8-FL1	488	PRT
78-24-3-A8-FL1	489	PRT
78-24-3-H4-FL2	490	PRT
78-25-1-F11-FL1	491	PRT
78-26-1-B5-FL1	492	PRT
78-27-3-D1-FL1	493	PRT
78-29-1-B2-FL1	494	PRT
78-29-4-B6-FL1	495	PRT
14-1-3-E6-FL1	496	PRT
30-9-1-G8-FL2	497	PRT
33-10-4-H2-FL2	498	PRT
33-10-4-H2-FL1	499	PRT
74-10-3-C9-FL2	500	PRT
33-97-4-G8-FL3	501	PRT
33-97-4-G8-FL2	502	PRT
33-104-4-H4-FL1	503	PRT
47-2-3-B3-FL1	504	PRT
47-37-4-G11-FL1	505	PRT
57-25-1-F10-FL2	506	PRT
58-19-3-D3-FL1	507	PRT
58-34-3-C9-FL2	508	PRT
58-48-4-E2-FL2	509	PRT
76-21-1-C4-FL1	510	PRT
78-26-2-H7-FL1	511	PRT
77-20-2-E11-FL1	512	PRT
47-1-3-F7-FL2	513	PRT

-136-

TABLE VIII

ID	Locations	PROSITE Signature Name
195	110-121	Aldehyde dehydrogenases csyteine active site
221	28-37	ATP synthase alpha and beta subunits signature
223	171-181	Regulator of chromosome condensation (RCC1) signature 2
225	90-112	Phosphatidylethanolamine-binding protein family signature
226	10-34	Protein kinases ATP-binding region signature

-137-

WHAT IS CLAIMED IS:

1.   A purified or isolated nucleic acid comprising the sequence of one of SEQ ID NOs: 40-140 and 242-377 or a sequence complementary thereto.

2.   A purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of
5   SEQ ID NOs: 40-140 and 242-377 or one of the sequences complementary thereto.

3.   A purified or isolated nucleic acid comprising the full coding sequences of one of SEQ ID NOs: 40, 42-44, 46, 48, 49, 51, 53, 60, 62-72, 76-78, 80-83, 85-88, 90, 93, 94, 97, 99-102, 104, 107-125, 127, 132, 135-138, 140 and 242-377wherein the full coding sequence comprises the sequence encoding signal peptide and the sequence encoding mature protein.

10   4.   A purified or isolated nucleic acid comprising the nucleotides of one of SEQ ID NOs: 40-44, 46, 48, 49, 51-53, 55, 56, 58-72, 75-78, 80-88, 90, 93, 94, 97, 99-125, 127, 132, 133, 135-138, 140, and 242-377 which encode a mature protein.

5.   A purified or isolated nucleic acid comprising the nucleotides of one of SEQ ID NOs: 40, 42-46, 48, 49, 51, 53, 57, 60, 62-73, 76-78, 80-83, 85-88, 90, 93-95, 97, 99-102, 104, 107-125, 127, 128, 130, 132, 134-140
15   and 242-377 which encode the signal peptide.

6.   A purified or isolated nucleic acid encoding a polypeptide having the sequence of one of the sequences of SEQ ID NOs: 141-241 and 378-513.

7.   A purified or isolated nucleic acid encoding a polypeptide having the sequence of a mature protein included in one of the sequences of SEQ ID NOs: 141-145, 147, 149, 150, 152-154, 156, 157, 159-172, 176-179, 181-
20   189, 191, 194, 195, 198, 200-226, 228, 233, 234, 236-239, 241 and 378-513.

8.   A purified or isolated nucleic acid encoding a polypeptide having the sequence of a signal peptide included in one of the sequences of SEQ ID NOs: 141, 143-147, 149, 150, 152, 154, 158, 161, 163-174, 177-179, 181-184, 186-189, 191, 194-196, 198, 200-203, 205, 208-226, 228, 229, 231, 233, 235-241, and 378-513.

9.   A purified or isolated protein comprising the sequence of one of SEQ ID NOs: 141-241 and 378-513.

25   10.   A purified or isolated polypeptide comprising at least 10 consecutive amino acids of one of the sequences of SEQ ID NOs: 141-241 and 378-513.

11.   An isolated or purified polypeptide comprising a signal peptide of one of the polypeptides of SEQ ID NOs: 141, 143-147, 149, 150, 152, 154, 158, 161, 163-174, 177-179, 181-184, 186-189, 191, 194-196, 198, 200-203, 205, 208-226, 228, 229, 231, 233, 235-241, and 378-513.

30   12.   An isolated or purified polypeptide comprising a mature protein of one of the polypeptides of SEQ ID NOs: 141-145, 147, 149, 150, 152-154, 156, 157, 159-172, 176-179, 181-189, 191, 194, 195, 198, 200-226, 228, 233, 234, 236-239, 241 and 378-513.

13.   A method of making a protein comprising one of the sequences of SEQ ID NO: 141-241 and 378-513, comprising the steps of:

·138·

obtaining a cDNA comprising one of the sequences of sequence of SEQ ID NO: 40-140 and 242-377;

inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter; and

introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said

cDNA.

5   14.   The method of Claim 13, further comprising the step of isolating said protein.

        15.   A protein obtainable by the method of Claim 14.

        16.   A host cell containing a recombinant nucleic acid of Claim 1.

        17.   A purified or isolated antibody capable of specifically binding to a protein having the sequence of one

of SEQ ID NOs: 141-241 and 378-513.

10   18.   In an array of polynucleotides of at least 15 nucleotides in length, the improvement comprising

inclusion in said array of at least one of the sequences of SEQ ID NOs: 40-140 and 242-377, or one of the sequences

complementary to the sequences of SEQ ID NOs: 40-140 and 242-377, or a fragment thereof of at least 15 consecutive

nucleotides.

        19.   A purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent

15 conditions to the sequence of one of SEQ ID NOs: 40-140 and 242-377 or a sequence complementary to one of the

sequences of SEQ ID NOs: 40-140 and 242-377.

        20.   A purified or isolated antibody capable of binding to a polypeptide comprising at least 10 consecutive

amino acids of the sequence of one of SEQ ID NOs: 141-241 and 378-513.
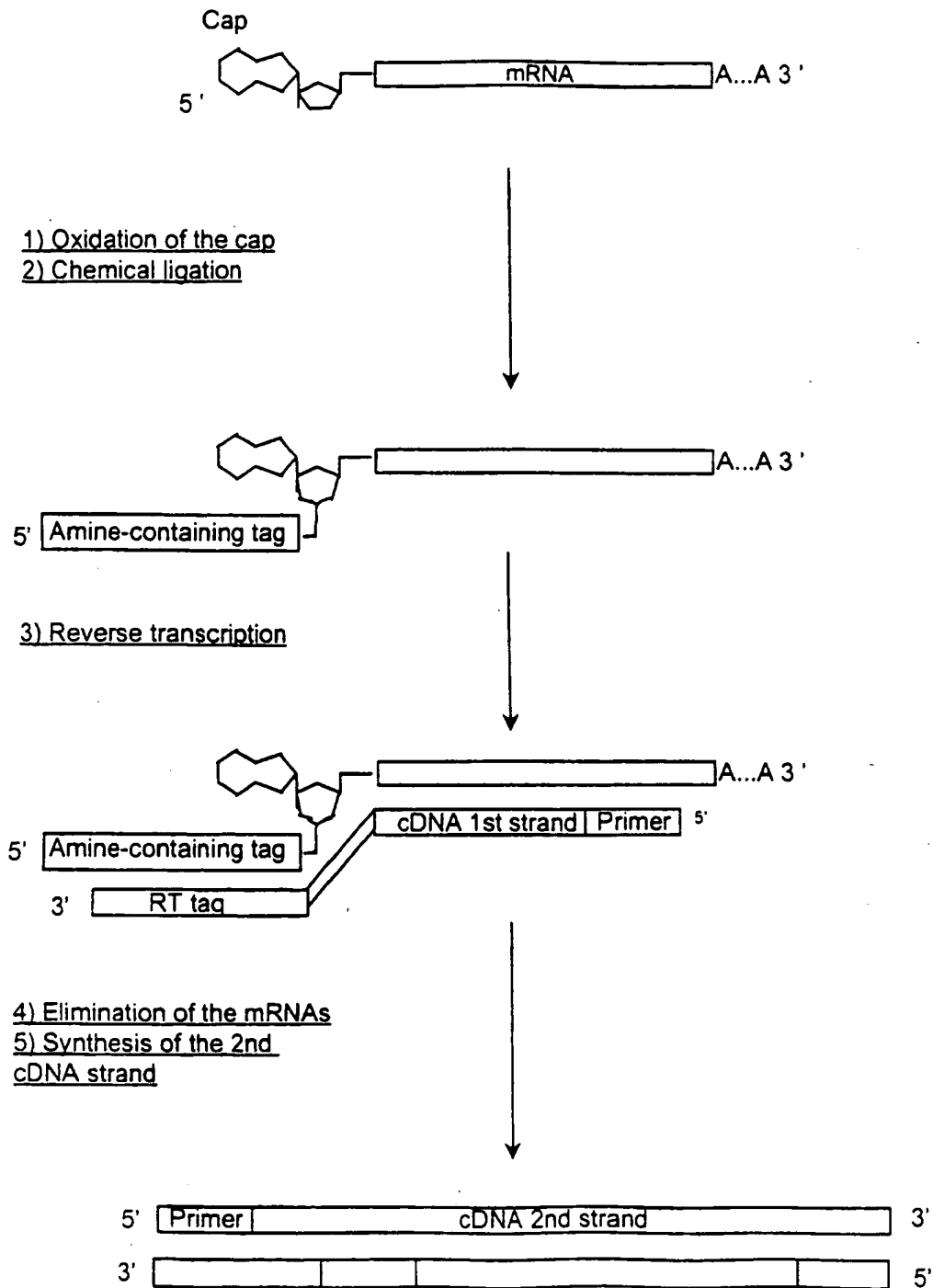
Figure 1

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3,5	0,121	0,036	0,467	0,664
4	0,096	0,06	0,519	0,708
4,5	0,078	0,079	0,565	0,745
5	0,062	0,098	0,615	0,782
5,5	0,05	0,127	0,659	0,813
6	0,04	0,163	0,694	0,836
6,5	0,033	0,202	0,725	0,855
7	0,025	0,248	0,763	0,878
7,5	0,021	0,304	0,78	0,889
8	0,015	0,368	0,816	0,909
8,5	0,012	0,418	0,836	0,92
9	0,009	0,512	0,856	0,93
9,5	0,007	0,581	0,863	0,934
10	0,006	0,679	0,835	0,919

FIGURE 2

FIGURE 3

4/15
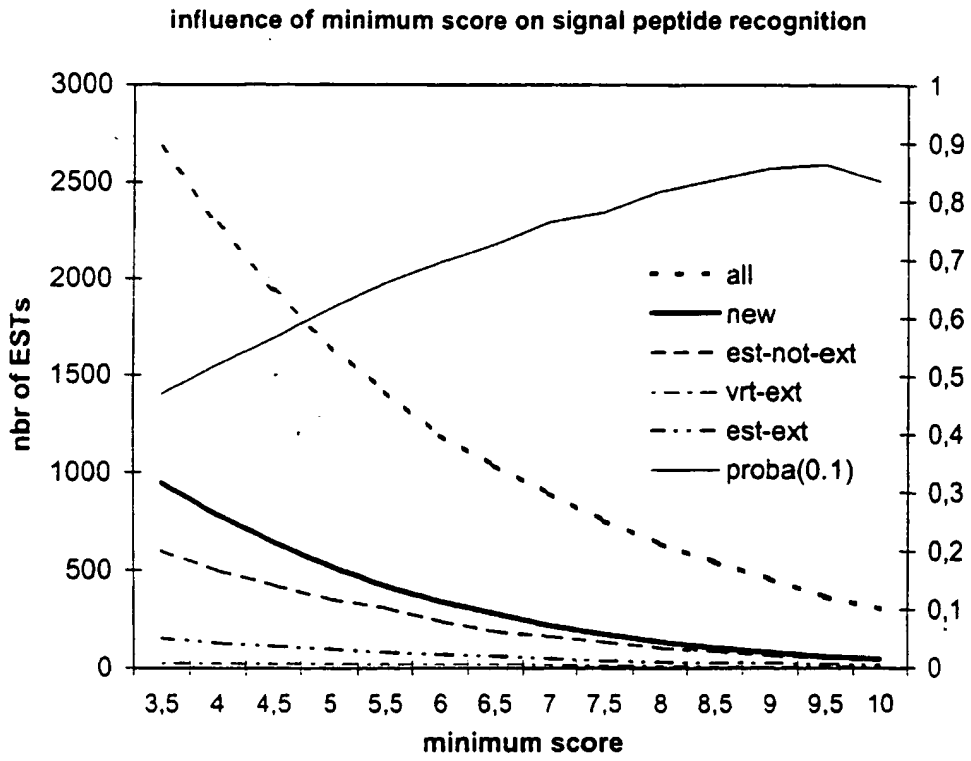
Minimum signal peptide score	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
3,5	2674	947	599	23	150
4	2278	784	499	23	126
4,5	1943	647	425	22	112
5	1657	523	353	21	96
5,5	1417	419	307	19	80
6	1190	340	238	18	68
6,5	1035	280	186	18	60
7	893	219	161	15	48
7,5	753	173	132	12	36
8	636	133	101	11	29
8,5	543	104	83	8	26
9	456	81	63	6	24
9,5	364	57	48	6	18
10	303	47	35	6	15

FIGURE 4

Tissue	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
Brain	329	131	75	3	24
Cancerous prostate	134	40	37	1	6
Cerebellum	17	9	1	0	6
Colon	21	11	4	0	0
Dystrophic muscle	41	18	8	0	1
Fetal brain	70	37	16	0	1
Fetal kidney	227	116	46	1	19
Fetal liver	13	7	2	0	0
Heart	30	15	7	0	1
Hypertrophic prostate	86	23	22	2	2
Kidney	10	7	3	0	0
Large intestine	21	8	4	0	1
Liver	23	9	6	0	0
Lung	24	12	4	0	1
Lung (cells)	57	38	6	0	4
Lymph ganglia	163	60	23	2	12
Lymphocytes	23	6	4	0	2
Muscle	33	16	6	0	4
Normal prostate	181	61	45	7	11
Ovary	90	57	12	1	2
Pancreas	48	11	6	0	1
Placenta	24	5	1	0	0
Prostate	34	16	4	0	2
Spleen	56	28	10	0	1
Substantia nigra	108	47	27	1	6
Surrenals	15	3	3	1	0
Testis	131	68	25	1	8
Thyroid	17	8	2	0	2
Umbilical cord	55	17	12	1	3
Uterus	28	15	3	0	2
Non tissue-specific	568	48	177	2	28
Total	2677	947	601	23	150

FIGURE 5
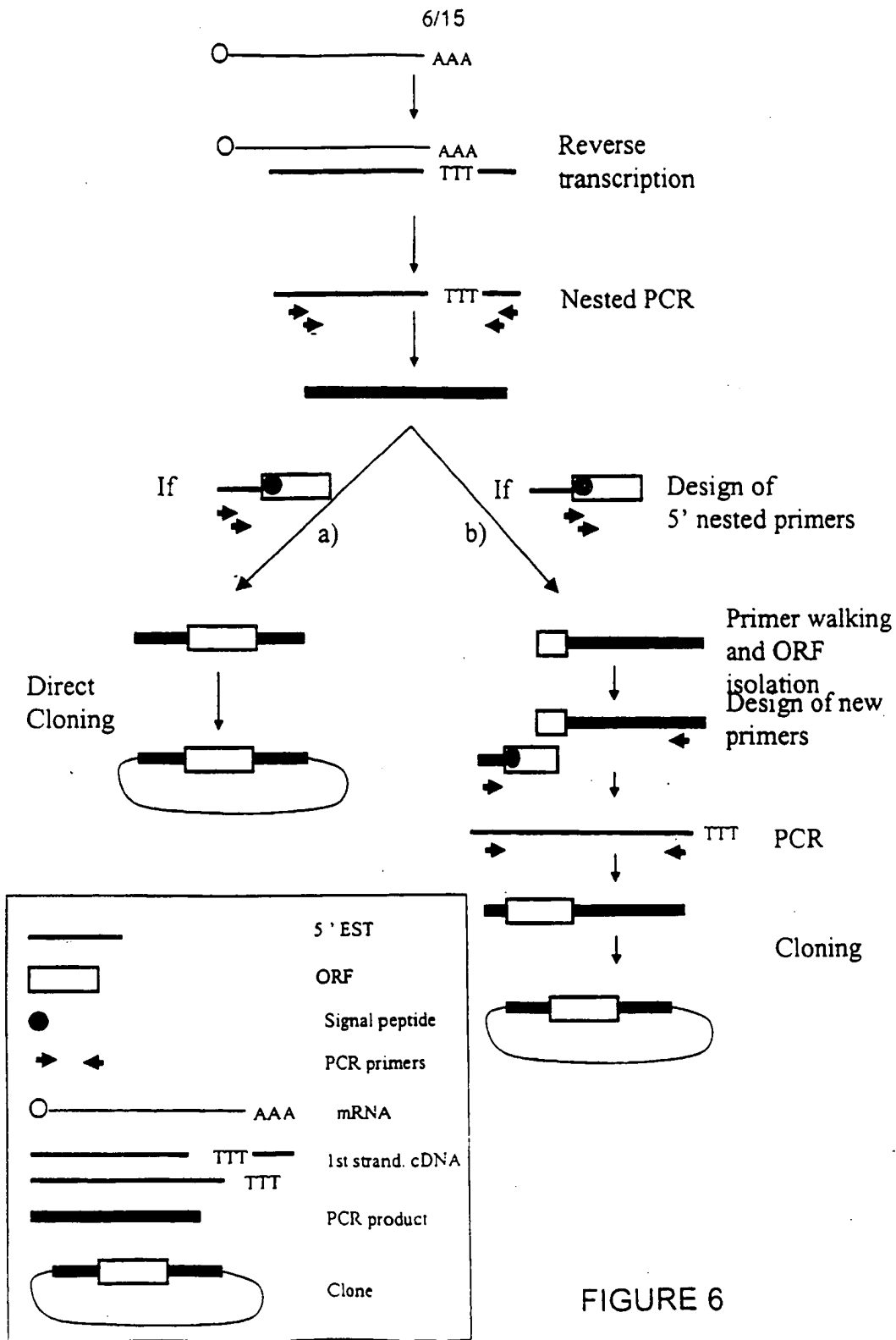
FIGURE 6
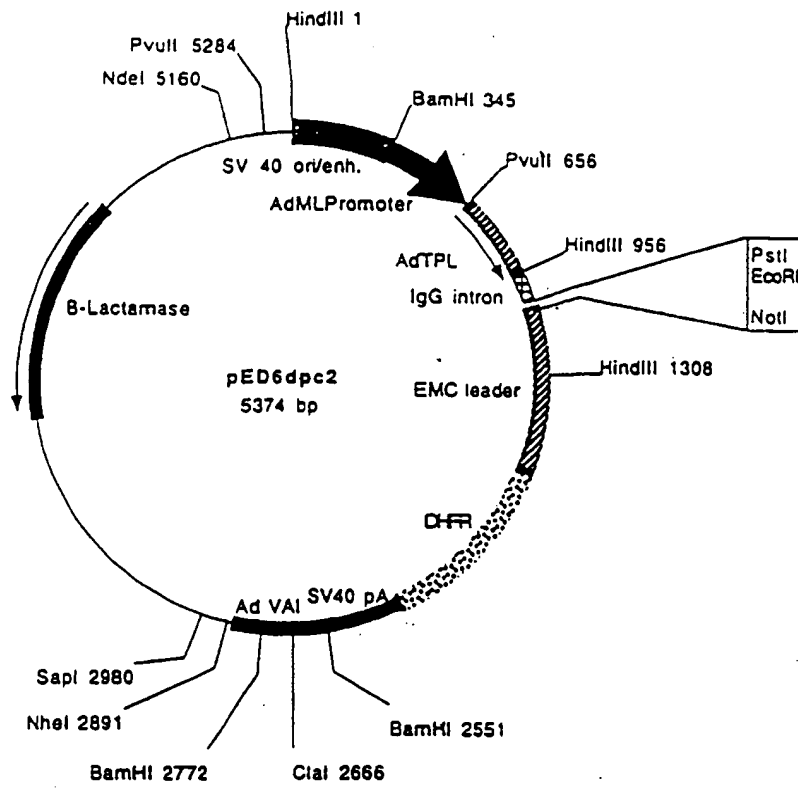
Plasmid name: pED6dpc2
Plasmid size: 5374 bp

FIGURE 7

Description of promoters structure isolated from SignalTag 5 'ESTs
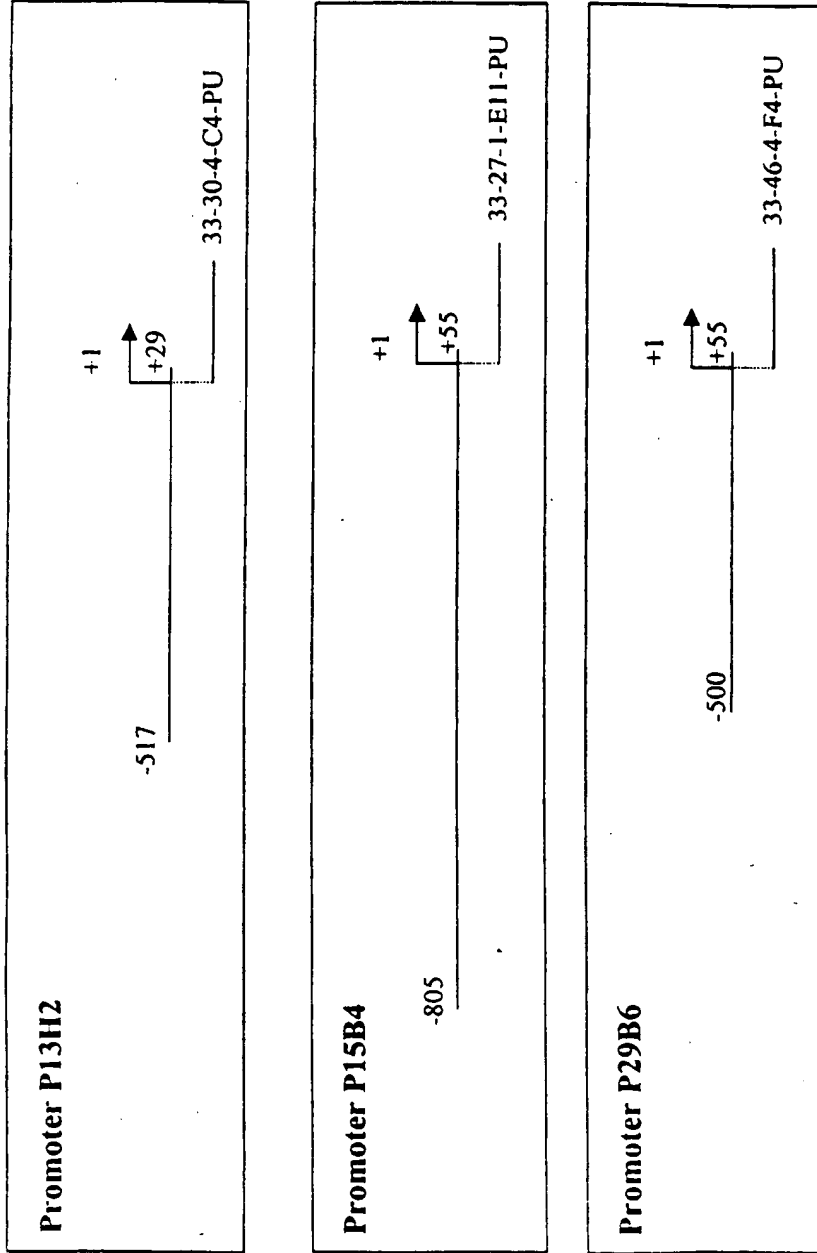
FIGURE 8

Description of Transcription Factor Binding Sites present on promoters
isolated from SignalTag sequences

Promoter sequence P13H2 (546 bp):

Matrix	Position	Orientation	Score	Length	Sequence
CMYB_01	-502	+	0.983	9	TGTCAGTTG
MYOD_Q6	-501	-	0.961	10	CCCAACTGAC
S8_01	-444	-	0.960	11	AATAGAATTAG
S8_01	-425	+	0.966	11	AACTAAATTAG
DELTAEF1_01	-390	-	0.960	11	GCACACCTCAG
GATA_C	-364	-	0.964	11	AGATAAATCCA
CMYB_01	-349	+	0.958	9	CTTCAGTTG
GATA1_02	-343	+	0.959	14	TTGTAGATAGGACA
GATA_C	-339	+	0.953	11	AGATAGGACAT
TAL1ALPHAE47_01	-235	+	0.973	16	CATAACAGATGGTAAG
TAL1BETAE47_01	-235	+	0.983	16	CATAACAGATGGTAAG
TAL1BETAITF2_01	-235	+	0.978	16	CATAACAGATGGTAAG
MYOD_Q6	-232	-	0.954	10	ACCATCTGTT
GATA1_04	-217	-	0.953	13	TCAAGATAAAGTA
IK1_01	-126	+	0.963	13	AGTTGGGAATTCC
IK2_01	-126	+	0.985	12	AGTTGGGAATTC
CREL_01	-123	+	0.962	10	TGGGAATTCC
GATA1_02	-96	+	0.950	14	TCAGTGATATGGCA
SRY_02	-41	-	0.951	12	TAAAACAAAACA
E2F_02	-33	+	0.957	8	TTTAGCGC
MZF1_01	-5	-	0.975	8	TGAGGGGA

Promoter sequence P15B4 (861bp) :

Matrix	Position	Orientation	Score	Length	Sequence
NFY_Q6	-748	-	0.956	11	GGACCAATCAT
MZF1_01	-738	+	0.962	8	CCTGGGGA
CMYB_01	-684	+	0.994	9	TGACCGTTG
VMYB_02	-682	-	0.985	9	TCCAACGGT
STAT_01	-673	+	0.968	9	TTCCTGGAA
STAT_01	-673	-	0.951	9	TTCCAGGAA
MZF1_01	-556	-	0.956	8	TTGGGGGA
IK2_01	-451	+	0.965	12	GAATGGGATTTC
MZF1_01	-424	+	0.986	8	AGAGGGGA
SRY_02	-398	-	0.955	12	GAAAACAAAACA
MZF1_01	-216	+	0.960	8	GAAGGGGA
MYOD_Q6	-190	+	0.981	10	AGCATCTGCC
DELTAEF1_01	-176	+	0.958	11	TCCCACCTTCC
S8_01	5	-	0.992	11	GAGGCAATTAT
MZF1_01	16	-	0.986	8	AGAGGGGA

Promoter sequence P29B6 (555 bp) :

Matrix	Position	Orientation	Score	Length	Sequence
ARNT_01	-311	+	0.964	16	GGACTCACGTGCTGCT
NMYC_01	-309	+	0.965	12	ACTCACGTGCTG
USF_01	-309	+	0.985	12	ACTCACGTGCTG
USF_01	-309	-	0.985	12	CAGCACGTGAGT
NMYC_01	-309	-	0.956	12	CAGCACGTGAGT
MYCMAX_02	-309	-	0.972	12	CAGCACGTGAGT
USF_C	-307	+	0.997	8	TCACGTGC
USF_C	-307	-	0.991	8	GCACGTGA
MZF1_01	-292	-	0.968	8	CATGGGGA
ELK1_02	-105	+	0.963	14	CTCTCCGGAAGCCT
CETS1P54_01	-102	+	0.974	10	TCCGGAAGCC
AP1_Q4	-42	-	0.963	11	AGTGACTGAAC
AP1FJ_Q2	-42	-	0.961	11	AGTGACTGAAC
PADS_C	45	+	1.000	9	TGTGGTCTC

Figure 9

100.0% identity in 125 aa overlap

```
                   10        20        30        40        50        60
SEQ ID NO: 217 MADEELEALRRQRLAELQAKHGDPGDAAQQEAKHREAEMRNSILAQVLDQSARARLSNLA
               X::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SEQ ID NO: 516 MADEELEALRRQRLAELQAKHGDPGDAAQQEAKHREAEMRNSILAQVLDQSARARLSNLA
                   10        20        30        40        50        60


                   70        80        90       100       110       120
SEQ ID NO: 217 LVKPEKTKAVENYLIQMARYGQLSEKVSEQGLIEILKKVSQQTEKTTTVKFNRRKVMDSD
               ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SEQ ID NO: 516 LVKPEKTKAVENYLIQMARYGQLSEKVSEQGLIEILKKVSQQTEKTTTVKFNRRKVMDSD
                   70        80        90       100       110       120
```

```
SEQ ID NO: 217 EDDDY
               ::::X
SEQ ID NO: 516 EDDDY
```

FIGURE 10

CLUSTAL W(1.5) multiple sequence alignment


```
SEQ ID NO: 517     MFCPLKLILLPVLLDYSLGLNDLNVSPPELTVHVGDSALMGCVFQSTEDKCIFKIDWTLS
SEQ ID NO: 232     ------------------------------------MGCVFQSTEDKCIFKIDWTLS
SEQ ID NO: 174     ------------------------------------MGCVFQSTEDKRIFKIDWTLS
SEQ ID NO: 175     ------------------------------------MGCVFQSTVDKCIFKIDWTLS
                                                       ******** ** *********


SEQ ID NO: 517     PGEHAKDEYVLYYYSNLSVPIGRFQNRVHLMGDNLCNDGSLLLQDVQDVE----------
SEQ ID NO: 232     PGEHAKDEYVLYYYSNLSVPIGRFQNRVHLMGDILCNDGSLLLQDVQEADQGTYICEIRL
SEQ ID NO: 174     PGEHAKDEYVLYYYSNLSVPIGRFQNRVHLMGDNLCNDGSLLLQDVQEADQGTYICEIRL
SEQ ID NO: 175     PGEHAKDEYVLYY SNLSVPIGRFQNRVHLMGDILCNDGSLLLQDVQEADQGTYICEIRL
                   ******************************** ***********.  .


SEQ ID NO: 517     ------------------------------------------------------------
SEQ ID NO: 232     KGESQVFKKAVVLHVLPEEPKGTQMLT---------------------------------
SEQ ID NO: 174     KGESQVFKKAVVLHVLPEEPKELMVHVGGLIQMGCVFQSTEVKHVTKVEWIFSGRRAKEE
SEQ ID NO: 175     KGESQVFKKAVVLHVLPEEPKELMVHVGGLIQMGCVFQSTEVKHVTKVEWIFSGR--RAK


SEQ ID NO: 517     ------------------------------------------------------------
SEQ ID NO: 232     ------------------------------------------------------------
SEQ ID NO: 174     IVFRYYHKLRMSAEYSQSWGHFQNRVNLVGDIFRNDGSIMLQGVRESDGGNYTCSIHLGN
SEQ ID NO: 175     VTRRKHHCVREGSG----------------------------------------------


SEQ ID NO: 517     ------------------------------------------------------------
SEQ ID NO: 232     ------------------------------------------------------------
SEQ ID NO: 174     LVFKKTIVLHVSPEEPRTLVTPAALRPLVLGGNQLVIIVGIVCATILLLPVLILIVKKTC
SEQ ID NO: 175     ------------------------------------------------------------


SEQ ID NO: 517     --------------------
SEQ ID NO: 232     --------------------
SEQ ID NO: 174     GNKSSVNSTVLVKNTKKTNP
SEQ ID NO: 175     --------------------
```

FIGURE 11

99.6% identity in 225 aa overlap

```
                10        20        30        40        50        60
SEQ ID NO: 515 PTAVQKEEARQDVEALLSRTVRTQILTGKELRVATQEKEGSSGRCMLTLLGLSFILAGLI
                                             ::::::::::::::::::::::::::::::
SEQ ID NO: 231                              LRVATQEKEGSSGRCMLTLLGLSFILAGLI
                                             10        20        30


                70        80        90       100       110       120
SEQ ID NO: 515 VGGACIYKYFMPKSTIYRGEMCFFDSEDPANSLRGGEPNFLPVTEEADIREDDNIAIIDV
               ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SEQ ID NO: 231 VGGACIYKYFMPKSTIYRGEMCFFDSEDPANSLRGGEPNFLPVTEEADIREDDNIAIIDV
                        40        50        60        70        80        90


               130       140       150       160       170       180
SEQ ID NO: 515 PVPSFSDSDPAAIIHDFEKGMTAYLDLLLGNCYLMPLNTSIVMPPKNLVELFGKLASGRY
               :::::::::::::::::::::::::::::::::  :::::::::::::::::::::::::::::
SEQ ID NO: 231 PVPSFSDSDPAAIIHDFEKGMTAYLDLLLGICYLMPLNTSIVMPPKNLVELFGKLASGRY
                        100       110       120       130       140       150


               190       200       210       220       230       240
SEQ ID NO: 515 LPQTYVVREDLVAVEEIRDVSNLGIFIYQLCNNRKSFRLRRRDLLLGFNKRAIDKCWKIR
               ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SEQ ID NO: 231 LPQTYVVREDLVAVEEIRDVSNLGIFIYQLCNNRKSFRLRRRDLLLGFNKRAIDKCWKIR
                        160       170       180       190       200       210


               250       260
SEQ ID NO: 515 HFPNEFIVETKICQE
               :::::::::::::::
SEQ ID NO: 231 HFPNEFIVETKICQE
                        220
```

FIGURE 12

99.7% identity in 353 aa overlap

```
                                          10        20        30
SEQ ID NO:196                      MERGLKSADPRDGTGYTGWAGIAVLYLHLY
                                   :::::::::::::::::::::::::::::::
SEQ ID NO:518 LAEGYFDAAGRLTPEFSQRLTNKIRELLQQMERGLKSADPRDGTGYTGWAGIAVLYLHLY
                20        30        40        50        60        70


                40        50        60        70        80        90
SEQ ID NO:196 DVFGDPAYLQLAHGYVKQSLNCLTKRSITFLCGDAGPLAVAAVLYHKMNNEKQAEDCITR
              ::::::::::::::::::::::::::::::: :::::::::::::::::::::::::::::::
SEQ ID NO:518 DVFGDPAYLQLAHGYVKQSLNCLTKRSITPLCGDAGPLAVAAVLYHKMNNEKQAEDCITR
                80        90        100       110       120       130


                100       110       120       130       140       150
SEQ ID NO:196 LIHLNKIDPHAPNEMLYGRIGYIYALLFVNKNFGVEKTPQSHIQQICETILTSGENLARK
              :::::::::::::::::::::::::::::::::::::::::::: :::::::::::::::::
SEQ ID NO:518 LIHLNKIDPHAPNEMLYGRIGYIYALLFVNKNFGVEKIPQSHIQQICETILTSGENLARK
                140       150       160       170       180       190


               160       170       180       190       200       210
SEQ ID NO:196 RNFTAKSPLMYEWYQEYYVGAAHGLAGIYYYLMQPSLQVSQGKLHSLVKPSVDYVCQLKF
              ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SEQ ID NO:518 RNFTAKSPLMYEWYQEYYVGAAHGLAGIYYYLMQPSLQVSQGKLHSLVKPSVDYVCQLKF
                200      210       220       230       240       250


               220       230       240       250       260       270
SEQ ID NO:196 PSGNYPPCIGDNRDLLVHWCHGAPGVIYMLIQAYKVFREEKYLCDAYQCADVIWQYGLLK
              ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SEQ ID NO:518 PSGNYPPCIGDNRDLLVHWCHGAPGVIYMLIQAYKVFREEKYLCDAYQCADVIWQYGLLK
                260       270       280       290       300       310


               280       290       300       310       320       330
SEQ ID NO:196 KGYGLCHGSAGNAYAFLTLYNLTQDMKYLYRACKFAEWCLEYGEHGCRTPDTPFSLFEGM
              ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SEQ ID NO:518 KGYGLCHGSAGNAYAFLTLYNLTQDMKYLYRACKFAEWCLEYGEHGCRTPDTPFSLFEGM
                320       330       340       350       360       370


               340       350
SEQ ID NO:196 AGTIYFLADLLVPTKARFPAFEL
              :::::::::::::::::::::::::
SEQ ID NO:518 AGTIYFLADLLVPTKARFPAFEL
                380       390
```

FIGURE 13

98.5% identity in 194 aa overlap

```
                  90        100       110       120       130       140
SEQ ID NO:519 ARNLPPLTDAQKNKLRHLSVVTLAAKVKCIPYAVLLEALALRNVRQLEDLVIEAVYADVL
                  ::::::::.:::::::::::::::::::::::::::::::::::::::::::::::::::
SEQ ID NO:158 ARNLPPLTEAQKNKLRHLSVVTLAAKVKCIPYAVLLEALALRNVRQLEDLVIEAVYADVL
                  60        70        80        90       100       110


                 150       160       170       180       190       200
SEQ ID NO:519 RGSLDQRNQRLEVDYSIGRDIQRQDLSAIAQTLQEWCVGCEVVLSGIEEQVSRANQHKEQ
                  ::::::::::::::::::::::::::::::::::..:::::::::::::::::::::::::
SEQ ID NO:158 RGSLDQRNQRLEVDYSIGRDIQRQDLSAIARTLQEWCVGCEVVLSGIEEQVSRANQHKEQ
                 120       130       140       150       160       170


                 210       220       230       240       250       260
SEQ ID NO:519 QLGLKQQIESEVANLKKTIKVTTAAAAAATSQDPEQHLTELREPASGTNQRQPSKKASKG
                  :::::::::::::::::::::::::::::::::::::::::.::::::::::::::::::
SEQ ID NO:158 QLGLKQQIESEVANLKKTIKVTTAAAAAATSQDPEQHLTELREPAPGTNQRQPSKKASKG
                 180       190       200       210       220       230


                 270
SEQ ID NO:519 KGLRGSAKIWSKSN
                  ::::::::::::::
SEQ ID NO:158 KGLRGSAKIWSKSN
                 240       250
```

88.7% identity in 62 aa overlap

```
                  10        20        30        40        50        60
SEQ ID NO:519 MSAEVKVTGQNQEQFLLLAKSAKGAALATLIHQVLEAPGVYVFGELLDMPNVRELAESDF
                  :::::::::::::::::::::::::::::::::::::::::::::::::::::::.....
SEQ ID NO:158 MSAEVKVTGQNQEQFLLLAKSAKGAALATLIHQVLEAPGVYVFGELLDMPNVRELXARNL
                  10        20        30        40        50        60


SEQ ID NO:519 AS
                .X
SEQ ID NO:158 PP
```

FIGURE 14

68.9% identity in 74 aa overlap

```
                          10        20        30        40        50
SEQ ID NO:226  MIARRNPVPLRFLPDEARSLPPPKLTDPRLLYIGFLGYCSGLIDNLIRRRPIATAGLHR
               :..:..........:::::::::::::::::::::  ..::::::::::::::  :::::.   :::::
SEQ ID NO:514  MMTGRQGRATFQFLPDEARSLPPPKLTDPRLAFVGFLGYCSGLIDNAIRRRPVLLAGLHR
                          10        20        30        40      .  50        60


                  60        70
SEQ ID NO:226  QLLYITAFFLLDIIL
               ::::::!.:  ...  :
SEQ ID NO:514  QLLYITSFVFVGYYLLKRQDYMYAVRDHDMFSYIKSHPEDFPEKDKKTYGEVFEEFHPVR
                          70        80        90     .  100       110       120
```

FIGURE 15

<110> Dumas Milne Edwards, Jean-Baptiste
      Duclert, Aymeric
      Bougueleret, Lydie

<120> Extended cDNAS for Secreted Proteins

<130> GENSET.019A

<160> 519

<170> Patent.pm

<210> 1
<211> 47
<212> RNA
<213> Artificial Sequence

<220>
<221> In vitro transcription product
<221> modified_base
<222> (1)...(1)
<223> m7g

<400> 1
ngcauccuac ucccauccaa uuccacccua acuccuccca ucuccac   47

<210> 2
<211> 46
<212> RNA
<213> Artificial Sequence

<220>
<223> In vitro transcription product

<400> 2
gcauccuacu cccauccaau uccacccuaa cuccucccau cuccac   46

<210> 3
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> In vitro transcription product

<400> 3
atcaagaatt cgcacgagac catta   25

<210> 4
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 4
taatggtctc gtgcgaattc ttgat   25

<210> 5
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 5
ccgacaagac caacgtcaag gccgc   .   25

<210> 6
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 6
tcaccagcag gcagtggctt aggag   25

<210> 7
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 7
agtgattcct gctactttgg atggc   25

<210> 8
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 8
gcttggtctt gttctggagt ttaga   25

<210> 9

```
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 9
tccagaatgg gagacaagcc aattt                                    25



<210> 10
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 10
agggaggagg aaacagcgtg agtcc                                    25



<210> 11
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 11
atgggaaagg aaaagactca tatca                                    25



<210> 12
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 12
agcagcaaca atcaggacag cacag                                    25



<210> 13
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 13
atcaagaatt cgcacgagac catta                                    25
```

```
<210> 14
<211> 67
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 14
atcgttgaga ctcgtaccag cagagtcacg agagagacta cacggtactg gtttttttttt     60
tttttvn                                                                 67



<210> 15
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 15
ccagcagagt cacgagagag actacacgg                                        29



<210> 16
<211> 25
<212> DNA
<213> Artificial Sequence
<220>
<223> Oligonucleotide

<400> 16
cacgagagag actacacggt actgg                                           25



<210> 17
<211> 526
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> complement(261..376)
<223> blastn

<221> misc_feature
<222> complement(380..486)
<223> blastn

<221> misc_feature
<222> complement(110..145)
<223> blastn

<221> misc_feature
<222> complement(196..229)
<223> blastn
```

```
<221> sig_peptide
<222> 90..140
<223> Von Heijne matrix

<400> 17
aatatrarac agctacaata ttccagggcc artcacttgc catttctcat aacagcgtca      60
gagagaaaga actgactgar acgtttgag atg aag aaa gtt ctc ctc ctg atc      113
                                 Met Lys Lys Val Leu Leu Leu Ile
                                     -15                      -10
aca gcc atc ttg gca gtg gct gtw ggt ttc cca gtc tct caa gac cag      161
Thr Ala Ile Leu Ala Val Ala Val Gly Phe Pro Val Ser Gln Asp Gln
            -5                   1               5
gaa cga gaa aaa aga agt atc agt gac agc gat gaa tta gct tca ggr      209
Glu Arg Glu Lys Arg Ser Ile Ser Asp Ser Asp Glu Leu Ala Ser Gly
        10              15              20
wtt ttt gtg ttc cct tac cca tat cca ttt cgc cca ctt cca cca att      257
Xaa Phe Val Phe Pro Tyr Pro Tyr Pro Phe Arg Pro Leu Pro Pro Ile
    25              30              35
cca ttt cca aga ttt cca tgg ttt aga cgt aan ttt cct att cca ata      305
Pro Phe Pro Arg Phe Pro Trp Phe Arg Arg Xaa Phe Pro Ile Pro Ile
40              45              50                      55
cct gaa tct gcc cct aca act ccc ctt cct agc gaa aag taaacaaraa      354
Pro Glu Ser Ala Pro Thr Thr Pro Leu Pro Ser Glu Lys
            60                  65
ggaaaagtca crataaacct ggtcacctga aattgaaatt gagccacttc cttgaaraat     414
caaaattcct gttaataaaa raaaaacaaa tgtaattgaa atagcacaca gcattctcta     474
gtcaatatct ttagtgatct tctttaataa acatgaaagc aaaaaaaaaa aa            526
```

```
<210> 18
<211> 17
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> 1..17
<223> Von Heijne matrix
      score 8.2
      seq LLLITAILAVAVG/FP

<400> 18
Met Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val
1               5                   10                  15
Gly
```

```
<210> 19
<211> 822
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 260..464
<223> blastn

<221> misc_feature
<222> 118..184
```

```
<223> blastn

<221> misc_feature
<222> 56..113
<223> blastn

<221> misc_feature
<222> 454..485
<223> blastn

<221> misc_feature
<222> 118..545
<223> blastn

<221> misc_feature
<222> 65..369
<223> blastn

<221> misc_feature
<222> 61..399
<223> blastn

<221> misc_feature
<222> 408..458
<223> blastn

<221> misc_feature
<222> 60..399
<223> blastn

<221> misc_feature
<222> 393..432
<223> blastn

<221> sig_peptide
<222> 346..408
<223> Von Heijne matrix

<400> 19
actcctttta gcatagggge ttcggcgcca gcggccagcg ctagtcggtc tggtaagtgc     60
ctgatgccga gttccgtctc tcgcgtcttt tcctggtccc aggcaaagcg gasgnagatc    120
ctcaaacggc ctagtgcttc gcgcttccgg agaaaatcag cggtctaatt aattcctctg    180
gtttgttgaa gcagttacca agaatcttca acccttttccc acaaaagcta attgagtaca    240
cgttcctgtt gagtacacgt tcctgttgat ttacaaaagg tgcaggtatg agcaggtctg    300
aagactaaca ttttgtgaag ttgtaaaaca gaaaacctgt tagaa atg tgg tgg ttt    357
                                                  Met Trp Trp Phe
                                                  -20

cag caa ggc ctc agt ttc ctt cct tca gcc ctt gta att tgg aca tct    405
Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val Ile Trp Thr Ser
        -15                 -10                 -5

gct gct ttc ata ttt tca tac att act gca gta aca ctc cac cat ata    453
Ala Ala Phe Ile Phe Ser Tyr Ile Thr Ala Val Thr Leu His His Ile
    1               5                   10                  15

gac ccg gct tta cct tat atc agt gac act ggt aca gta gct cca raa    501
Asp Pro Ala Leu Pro Tyr Ile Ser Asp Thr Gly Thr Val Ala Pro Xaa
                20                  25                  30

aaa tgc tta ttt ggg gca atg cta aat att gcg gca gtt tta tgt caa    549
Lys Cys Leu Phe Gly Ala Met Leu Asn Ile Ala Ala Val Leu Cys Gln
            35                  40                  45

aaa tagaaatcag gaarataatt caacttaaag aakttcattt catgaccaaa          602
Lys

ctcttcaraa acatgtcttt acaagcatat ctcttgtatt gctttctaca ctgttgaatt    662
```

```
gtctggcaat atttctgcag tggaaaattt gatttarmta gttcttgact gataaatatg   722
gtaaggtggg cttttccccc tgtgtaattg gctactatgt cttactgagc caagttgtaw   782
tttgaaataa aatgatatga gagtgacaca aaaaaaaaaa                         822
```

```
<210> 20
<211> 21
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> 1..21
<223> Von Heijne matrix
      score 5.5
      seq SFLPSALVIWTSA/AF

<400> 20
Met Trp Trp Phe Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val
1               5                   10                  15
Ile Trp Thr Ser Ala
            20
```

```
<210> 21
<211> 405
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> complement(103..398)
<223> blastn

<221> sig_peptide
<222> 185..295
<223> Von Heijne matrix

<400> 21
atcaccttct tctccatcct tstctgggcc agtccccarc ccagtccctc tcctgacctg   60
cccagcccaa gtcagccttc agcacgcgct tttctgcaca cagatattcc aggcctacct   120
ggcattccag gacctccgma atgatgctcc agtcccttac aagcgcttcc tggatgaggg   180
tggc atg gtg ctg acc acc ctc ccc ttg ccc tct gcc aac agc cct gtg   229
     Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val
         -35                 -30                 -25
aac atg ccc acc act ggc ccc aac agc ctg agt tat gct agc tct gcc   277
Asn Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala
    -20                 -15                 -10
ctg tcc ccc tgt ctg acc gct cca aak tcc ccc cgg ctt gct atg atg   325
Leu Ser Pro Cys Leu Thr Ala Pro Xaa Ser Pro Arg Leu Ala Met Met
    -5              1               5                   10
cct gac aac taaatatcct tatccaaatc aataaarwra raatcctccc           374
Pro Asp Asn
tccaraaggg tttctaaaaa caaaaaaaaa a                                 405
```

```
<210> 22
<211> 37
<212> PRT
```

<213> Homo sapiens

<220>
<221> SIGNAL
<222> 1..37
<223> Von Heijne matrix
     score 5.9
     seq LSYASSALSPCLT/AP

<400> 22
Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val Asn
1   5   10   15
Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala Leu
            20   25   30
Ser Pro Cys Leu Thr
            35

<210> 23
<211> 496
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 149..331
<223> blastn

<221> misc_feature
<222> 328..485
<223> blastn

<221> misc_feature
<222> complement(182..496)
<223> blastn

<221> sig_peptide
<222> 196..240
<223> Von Heijne matrix

<400> 23
aaaaaattgg tcccagtttt caccctgccg cagggctggc tggggagggc agcggtttag   60
attagccgtg gcctaggccg tttaacgggg tgacacgagc ntgcagggcc gagtccaagg   120
cccggagata ggaccaaccg tcaggaatgc gaggaatgtt tttcttcgga ctctatcgag   180
gcacacagac agacc atg ggg att ctg tct aca gtg aca gcc tta aca ttt   231
                  Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe
                  -15   -10   -5
gcc ara gcc ctg gac ggc tgc aga aat ggc att gcc cac cct gca agt   279
Ala Xaa Ala Leu Asp Gly Cys Arg Asn Gly Ile Ala His Pro Ala Ser
            1   5   10
gag aag cac aga ctc gag aaa tgt agg gaa ctc gag asc asc cac tcg   327
Glu Lys His Arg Leu Glu Lys Cys Arg Glu Leu Glu Xaa Xaa His Ser
        15   20   25
gcc cca gga tca acc cas cac cga aga aaa aca acc aga aga aat tat   375
Ala Pro Gly Ser Thr Xaa His Arg Arg Lys Thr Thr Arg Arg Asn Tyr
30   35   40   45
tct tca gcc tgaaatgaak ccgggatcaa atggttgctg atcaragccc   424
Ser Ser Ala
atatttaaat tggaaaagtc aaattgasca ttattaaata aagcttgttt aatatgtctc   484
aaacaaaaaa aa   496

```
<210> 24
<211> 15
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> 1..15
<223> Von Heijne matrix
      score 5.5
      seq ILSTVTALTFAXA/LD

<400> 24
Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe Ala Xaa Ala
1               5                   10                  15



<210> 25
<211> 623
<212> DNA
<213> Homo sapiens

<220>
<221> sig_peptide
<222> 49..96
<223> Von Heijne matrix

<400> 25
aaagatccct gcagcccggc aggagagaag gctgagcctt ctggcgtc atg gag agg     57
                                                     Met Glu Arg
                                                             -15

ctc gtc cta acc ctg tgc acc ctc ccg ctg gct gtg gcg tct gct ggc     105
Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala Ser Ala Gly
            -10             -5                   1

tgc gcc acg acg cca gct cgc aac ctg agc tgc tac cag tgc ttc aag     153
Cys Ala Thr Thr Pro Ala Arg Asn Leu Ser Cys Tyr Gln Cys Phe Lys
        5               10                  15

gtc agc agc tgg acg gag tgc ccg ccc acc tgg tgc agc ccg ctg gac     201
Val Ser Ser Trp Thr Glu Cys Pro Pro Thr Trp Cys Ser Pro Leu Asp
20                  25                  30                  35

caa gtc tgc atc tcc aac gag gtg gtc gtc tct ttt aaa tgg agt gta     249
Gln Val Cys Ile Ser Asn Glu Val Val Val Ser Phe Lys Trp Ser Val
                40                  45                  50

cgc gtc ctg ctc agc aaa cgc tgt gct ccc aga tgt ccc aac gac aac     297
Arg Val Leu Leu Ser Lys Arg Cys Ala Pro Arg Cys Pro Asn Asp Asn
            55                  60                  65

atg aak ttc gaa tgg tcg ccg gcc ccc atg gtg caa ggc gtg atc acc     345
Met Xaa Phe Glu Trp Ser Pro Ala Pro Met Val Gln Gly Val Ile Thr
        70                  75                  80

agg cgc tgc tgt tcc tgg gct ctc tgc aac agg gca ctg acc cca cag     393
Arg Arg Cys Cys Ser Trp Ala Leu Cys Asn Arg Ala Leu Thr Pro Gln
        85                  90                  95

gag ggg cgc tgg gcc ctg cra ggg ggg ctc ctg ctc cag gac cct tcg     441
Glu Gly Arg Trp Ala Leu Xaa Gly Gly Leu Leu Leu Gln Asp Pro Ser
100                 105                 110                 115

agg ggc ara aaa acc tgg gtg cgg cca cag ctg ggg ctc cca ctc tgc     489
Arg Gly Xaa Lys Thr Trp Val Arg Pro Gln Leu Gly Leu Pro Leu Cys
                120                 125                 130

ctt ccc awt tcc aac ccc ctc tgc cca rgg gaa acc cag gaa gga         534
```

Leu Pro Xaa Ser Asn Pro Leu Cys Pro Xaa Glu Thr Gln Glu Gly
            135   140   145
taacactgtg ggtgccccca cctgtgcatt gggaccacra cttcaccctc ttggaracaa   594
taaactctca tgcccccaaa aaaaaaaaa   623

<210> 26
<211> 16
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> 1..16
<223> Von Heijne matrix
      score 10.1
      seq LVLTLCTLPLAVA/SA

<400> 26
Met Glu Arg Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala
1   5   10   15

<210> 27
<211> 848
<212> DNA
<213> Homo sapiens

<220>
<221> sig_peptide
<222> 32..73
<223> Von Heijne matrix

<400> 27
aactttgcct tgtgttttcc accctgaaag a atg ttg tgg ctg ctc ttt ttt   52
                                  Met Leu Trp Leu Leu Phe Phe
                                                    -10
ctg gtg act gcc att cat gct gaa ctc tgt caa cca ggt gca gaa aat   100
Leu Val Thr Ala Ile His Ala Glu Leu Cys Gln Pro Gly Ala Glu Asn
        -5   1   5
gct ttt aaa gtg aga ctt agt atc aga aca gct ctg gga gat aaa gca   148
Ala Phe Lys Val Arg Leu Ser Ile Arg Thr Ala Leu Gly Asp Lys Ala
10   15   20   25
tat gcc tgg gat acc aat gaa gaa tac ctc ttc aaa gcg atg gta gct   196
Tyr Ala Trp Asp Thr Asn Glu Glu Tyr Leu Phe Lys Ala Met Val Ala
                30   35   40
ttc tcc atg aga aaa gtt ccc aac aga gaa gca aca gaa att tcc cat   244
Phe Ser Met Arg Lys Val Pro Asn Arg Glu Ala Thr Glu Ile Ser His
                45   50   55
gtc cta ctt tgc aat gta acc cag agg gta tca ttc tgg ttt gtg gtt   292
Val Leu Leu Cys Asn Val Thr Gln Arg Val Ser Phe Trp Phe Val Val
                60   65   70
aca gac cct tca aaa aat cac acc ctt cct gct gtt gag gtg caa tca   340
Thr Asp Pro Ser Lys Asn His Thr Leu Pro Ala Val Glu Val Gln Ser
                75   80   85
gcc ata aga atg aac aag aac cgg atc aac aat gcc ttc ttt cta aat   388
Ala Ile Arg Met Asn Lys Asn Arg Ile Asn Asn Ala Phe Phe Leu Asn
90   95   100   105
gac caa act ctg gaa ttt tta aaa atc cct tcc aca ctt gca cca ccc   436
Asp Gln Thr Leu Glu Phe Leu Lys Ile Pro Ser Thr Leu Ala Pro Pro

```
                   110                    115                    120
atg gac cca tct gtg ccc atc tgg att att ata ttt ggt gtg ata ttt      484
Met Asp Pro Ser Val Pro Ile Trp Ile Ile Ile Phe Gly Val Ile Phe
              125                    130                    135
tgc atc atc ata gtt gca att gca cta ctg att tta tca ggg atc tgg      532
Cys Ile Ile Ile Val Ala Ile Ala Leu Leu Ile Leu Ser Gly Ile Trp
          140                    145                    150
caa cgt ada ara aag aac aaa gaa cca tct gaa gtg gat gac gct gaa      580
Gln Arg Xaa Xaa Lys Asn Lys Glu Pro Ser Glu Val Asp Asp Ala Glu
      155                    160                    165
rat aak tgt gaa aac atg atc aca att gaa aat ggc atc ccc tct gat      628
Xaa Xaa Cys Glu Asn Met Ile Thr Ile Glu Asn Gly Ile Pro Ser Asp
170                    175                    180                    185
ccc ctg gac atg aag gga ggg cat att aat gat gcc ttc atg aca gag      676
Pro Leu Asp Met Lys Gly Gly His Ile Asn Asp Ala Phe Met Thr Glu
              190                    195                    200
gat gag agg ctc acc cct ctc tgaagggctg ttgttctgct tcctcaaraa        727
Asp Glu Arg Leu Thr Pro Leu
              205
attaaacatt tgtttctgtg tgactgctga gcatcctgaa ataccaagag cagatcatat   787
wttttgtttc accattcttc ttttgtaata aattttgaat gtgcttgaaa aaaaaaaaa    847
c                                                                    848
```

```
<210> 28
<211> 14
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> 1..14
<223> Von Heijne matrix
      score 10.7
      seq LWLLFFLVTAIHA/EL

<400> 28
Met Leu Trp Leu Leu Phe Phe Leu Val Thr Ala Ile His Ala
1               5                   10



<210> 29
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 29
gggaagatgg agatagtatt gcctg                                          25



<210> 30
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
```

```
<223> Olignucleotide

<400> 30
ctgccatgta catgatagag agattc                                                      26




<210> 31
<211> 546
<212> DNA
<213> Homo sapiens

<220>
<221> promoter
<222> 1..517

<221> transcription start site
<222> 518

<221> protein_bind
<222> 17..25
<223> matinspector prediction
      name CMYB_01
      score 0.983
      sequence tgtcagttg

<221> protein_bind
<222> complement(18..27)
<223> matinspector prediction
      name MYOD_Q6   .
      score 0.961
      sequence cccaactgac

<221> protein_bind
<222> complement(75..85)
<223> matinspector prediction
      name S8_01
      score 0.960
      sequence aatagaattag

<221> protein_bind
<222> 94..104
<223> matinspector prediction
      name S8_01
      score 0.966
      sequence aactaaattag

<221> protein_bind
<222> complement(129..139)
<223> matinspector prediction
      name DELTAEF1_01
      score 0.960
      sequence gcacacctcag

<221> protein_bind
<222> complement(155..165)
<223> matinspector prediction
      name GATA_C
      score 0.964
      sequence agataaatcca

<221> protein_bind
```

```
<222>  170..178
<223>  matinspector prediction
       name CMYB_01
       score 0.958
       sequence cttcagttg

<221>  protein_bind
<222>  176..189
<223>  matinspector prediction
       name GATA1_02
       score 0.959
       sequence ttgtagataggaca

<221>  protein_bind
<222>  180..190
<223>  matinspector prediction
       name GATA_C
       score 0.953
       sequence agataggacat

<221>  protein_bind
<222>  284..299
<223>  matinspector prediction
       name TAL1ALPHAE47_01
       score 0.973
       sequence cataacagatggtaag

<221>  protein_bind
<222>  284..299
<223>  matinspector prediction
       name TAL1BETAE47_01
       score 0.983
       sequence cataacagatggtaag

<221>  protein_bind
<222>  284..299
<223>  matinspector prediction
       name TAL1BETAITF2_01
       score 0.978
       sequence cataacagatggtaag

<221>  protein_bind
<222>  complement(287..296)
<223>  matinspector prediction
       name MYOD_Q6
       score 0.954
       sequence accatctgtt

<221>  protein_bind
<222>  complement(302..314)
<223>  matinspector prediction
       name GATA1_04
       score 0.953
       sequence tcaagataaagta

<221>  protein_bind
<222>  393..405
<223>  matinspector prediction
       name IK1_01
       score 0.963
       sequence agttgggaattcc
```

```
<221> protein_bind
<222> 393..404
<223> matinspector prediction
      name IK2_01
      score 0.985
      sequence agttgggaattc


<221> protein_bind
<222> 396..405
<223> matinspector prediction
      name CREL_01
      score 0.962
      sequence tgggaattcc


<221> protein_bind
<222> 423..436
<223> matinspector prediction
      name GATA1_02
      score 0.950
      sequence tcagtgatatggca


<221> protein_bind
<222> complement(478..489)
<223> matinspector prediction
      name SRY_02
      score 0.951
      sequence taaaacaaaaca


<221> protein_bind
<222> 486..493
<223> matinspector prediction
      name E2F_02
      score 0.957
      sequence tttagcgc


<221> protein_bind
<222> complement(514..521)
<223> matinspector prediction
      name MZF1_01
      score 0.975
      sequence tgagggga


<400> 31
tgagtgcagt gttacatgtc agttgggtta agtttgttaa tgtcattcaa atcttctatg     60
tcttgatttg cctgctaatt ctattatttc tggaactaaa ttagtttgat ggttctatta    120
gttattgact gaggtgtgct aatctcccat tatgtggatt tatctatttc ttcagttgta    180
gataggacat tgatagatac ataagtacca ggacaaaagc agggagatct ttttttccaaa   240
atcaggagaa aaaaatgaca tctggaaaac ctatagggaa aggcataaca gatggtaagg    300
atactttatc ttgagtagga gagccttcct gtggcaacgt ggagaaggga agaggtcgta    360
gaattgagga gtcagctcag ttagaagcag ggagttggga attccgttca tgtgatttag    420
catcagtgat atggcaaatg tgggactaag ggtagtgatc agagggttaa aattgtgtgt    480
tttgttttag cgctgctggg gcatcgcctt gggtcccctc aaacagattc ccatgaatct    540
cttcat                                                               546
```

```
<210> 32
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
```

<223> Oligonucleotide

<400> 32
gtaccaggga ctgtgaccat tgc   23

<210> 33
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 33
ctgtgaccat tgctcccaag agag   24

<210> 34
<211> 861
<212> DNA
<213> Homo sapiens

<220>
<221> promoter
<222> 1..806

<221> transcription start site
<222> 807

<221> protein_bind
<222> complement(60..70)
<223> matinspector prediction
      name NFY_Q6
      score 0.956
      sequence ggaccaatcat

<221> protein_bind
<222> 70..77
<223> matinspector prediction
      name MZF1_01
      score 0.962
      sequence cctgggga

<221> protein_bind
<222> 124..132
<223> matinspector prediction
      name CMYB_01
      score 0.994
      sequence tgaccgttg

<221> protein_bind
<222> complement(126..134)
<223> matinspector prediction
      name VMYB_02
      score 0.985
      sequence tccaacggt

<221> protein_bind
<222> 135..143

```
<223> matinspector prediction
      name STAT_01
      score 0.968
      sequence ttcctggaa

<221> protein_bind
<222> complement(135..143)
<223> matinspector prediction
      name STAT_01
      score 0.951
      sequence ttccaggaa

<221> protein_bind
<222> complement(252..259)
<223> matinspector prediction
      name MZF1_01
      score 0.956
      sequence ttggggga

<221> protein_bind
<222> 357..368
<223> matinspector prediction
      name IK2_01
      score 0.965
      sequence gaatgggatttc

<221> protein_bind
<222> 384..391
<223> matinspector prediction
      name MZF1_01
      score 0.986
      sequence agagggga

<221> protein_bind
<222> complement(410..421)
<223> matinspector prediction
      name SRY_02
      score 0.955
      sequence gaaaacaaaaca

<221> protein_bind
<222> 592..599
<223> matinspector prediction
      name MZF1_01
      score 0.960
      sequence gaagggga

<221> protein_bind
<222> 618..627
<223> matinspector prediction
      name MYOD_Q6
      score 0.981
      sequence agcatctgcc

<221> protein_bind
<222> 632..642
<223> matinspector prediction
      name DELTAEF1_01
      score 0.958
      sequence tcccaccttcc

<221> protein_bind
```

<222> complement(813..823)
<223> matinspector prediction
    name S8_01
    score 0.992
    sequence gaggcaattat

<221> protein_bind
<222> complement(824..831)
<223> matinspector prediction
    name MZF1_01
    score 0.986
    sequence agagggga

<400> 34
tactataggg cacgcgtggt cgacggccgg gctgttctgg agcagagggc atgtcagtaa   60
tgattggtcc ctggggaagg tctggctggc tccagcacag tgaggcattt aggtatctct   120
cggtgaccgt tggattcctg gaagcagtag ctgttctgtt tggatctggt agggacaggg   180
ctcagagggc taggcacgag ggaaggtcag aggagaaggs aggsarggcc cagtgagarg   240
ggagcatgcc ttccccccaac cctggcttsc ycttggymam agggcgktty tgggmacttr   300
aaytcagggc ccaascagaa scacaggccc aktcntggct smaagcacaa tagcctgaat   360
gggatttcag gttagncagg gtgagagggg aggctctctg gcttagtttt gttttgtttt   420
ccaaatcaag gtaacttgct cccttctgct acgggccttg gtcttggctt gtcctcaccc   480
agtcggaact ccctaccact ttcaggagag tggttttagg cccgtggggc tgttctgttc   540
caagcagtgt gagaacatgg ctggtagagg ctctagctgt gtgcgggggcc tgaaggggag   600
tgggttctcg cccaaagagc atctgcccat ttcccaccectt cccttctccc accagaagct   660
tgcctgagct gtttggacaa aaatccaaac cccacttggc tactctggcc tggcttcagc   720
ttggaaccca atacctaggc ttacaggcca tcctgagcca ggggcctctg gaaattctct   780
tcctgatggt cctttaggtt tgggcacaaa atataattgc ctctccectc tcccattttc   840
tctcttggga gcaatggtca c   861

<210> 35
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 35
ctgggatgga aggcacggta   20

<210> 36
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 36
gagaccacac agctagacaa   20

<210> 37
<211> 555
<212> DNA
<213> Homo sapiens

```
<220>
<221> promoter
<222> 1..500

<221> transcription start site
<222> 501

<221> protein_bind
<222> 191..206
<223> matinspector prediction
      name ARNT_01
      score 0.964
      sequence ggactcacgtgctgct

<221> protein_bind
<222> 193..204
<223> matinspector prediction
      name NMYC_01
      score 0.965
      sequence actcacgtgctg

<221> protein_bind
<222> 193..204
<223> matinspector prediction
      name USF_01
      score 0.985
      sequence actcacgtgctg

<221> protein_bind
<222> complement(193..204)
<223> matinspector prediction
      name USF_01
      score 0.985
      sequence cagcacgtgagt

<221> protein_bind
<222> complement(193..204)
<223> matinspector prediction
      name NMYC_01
      score 0.956
      sequence cagcacgtgagt

<221> protein_bind
<222> complement(193..204)
<223> matinspector prediction
      name MYCMAX_02
      score 0.972
      sequence cagcacgtgagt

<221> protein_bind
<222> 195..202
<223> matinspector prediction
      name USF_C
      score 0.997
      sequence tcacgtgc

<221> protein_bind
<222> complement(195..202)
<223> matinspector prediction
      name USF_C
      score 0.991
```

```
      sequence gcacgtga

<221> protein_bind
<222> complement(210..217)
<223> matinspector prediction
      name MZF1_01
      score 0.968
      sequence catgggga

<221> protein_bind
<222> 397..410
<223> matinspector prediction
      name ELK1_02
      score 0.963
      sequence ctctccggaagcct

<221> protein_bind
<222> 400..409
<223> matinspector prediction
      name CETS1P54_01
      score 0.974
      sequence tccggaagcc

<221> protein_bind
<222> complement(460..470)
<223> matinspector prediction
      name AP1_Q4
      score 0.963
      sequence agtgactgaac

<221> protein_bind
<222> complement(460..470)
<223> matinspector prediction
      name AP1FJ_Q2
      score 0.961
      sequence agtgactgaac

<221> protein_bind
<222> 547..555
<223> matinspector prediction
      name PADS_C
      score 1.000
      sequence tgtggtctc

<400> 37
ctatagggca cgcktggtcg acggcccggg ctggtctggt ctgtkgtgga gtcgggttga      60
aggacagcat ttgtkacatc tggtctactg caccttccct ctgccgtgca cttggccttt     120
kawaagctca gcaccggtgc ccatcacagg gccggcagca cacacatccc attactcaga     180
aggaactgac ggactcacgt gctgctccgt ccccatgagc tcagtggacc tgtctatgta     240
gagcagtcag acagtgcctg ggatagagtg agagttcagc cagtaaatcc aagtgattgt     300
cattcctgtc tgcattagta actcccaacc tagatgtgaa aacttagttc tttctcatag     360
gttgctctgc ccatggtccc actgcagacc caggcactct ccggaagcct ggaaatcacc     420
cgtgtcttct gcctgctccc gctcacatcc cacacttgtg ttcagtcact gagttacaga     480
ttttgcctcc tcaatttctc ttgtcttagt cccatcctct gttcccctgg ccagtttgtc     540
tagctgtgtg gtctc                                                       555
```

<210> 38
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 38
ggccatacac ttgagtgac   19

<210> 39
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 39
atatagacaa acgcacacc   19

<210> 40
<211> 568
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 7..471

<221> sig_peptide
<222> 7..99
<223> Von Heijne matrix
      score 6.9
      seq LLLVPSALSLLLA/LL

<221> polyA_signal
<222> 537..542

<221> polyA_site
<222> 554..568

<400> 40
gggacc atg ttc acc agc acc ggc tcc agt ggg ctc tac aag gcg cct   48
       Met Phe Thr Ser Thr Gly Ser Ser Gly Leu Tyr Lys Ala Pro
          -30   -25   -20
ctg tcg aag agc ctt ctg ctg gtc ccc agt gcc ctc tcc ctc ctg ctc   96
Leu Ser Lys Ser Leu Leu Leu Val Pro Ser Ala Leu Ser Leu Leu Leu
      -15   -10   -5
gcc ctc ctc ctg cct cac tgc cag aag ccc ttt gtg tat gac ctt cac   144
Ala Leu Leu Leu Pro His Cys Gln Lys Pro Phe Val Tyr Asp Leu His
   1   5   10   15
gca gtc aag aac gac ttc cag att tgg agg ttg ata tgt gga aga ata   192
Ala Val Lys Asn Asp Phe Gln Ile Trp Arg Leu Ile Cys Gly Arg Ile
             20   25   30
att tgc ctt gat ttg aaa gat act ttc tgc agt agt ctg ctt att tat   240
Ile Cys Leu Asp Leu Lys Asp Thr Phe Cys Ser Ser Leu Leu Ile Tyr
          35   40   45
aat ttt agg ata ttt gaa aga aga tat gga agc aga aaa ttt gca tcc   288
Asn Phe Arg Ile Phe Glu Arg Arg Tyr Gly Ser Arg Lys Phe Ala Ser
       50   55   60

```
ttt ttg ctg ggt acc tgg gtt ttg tca gcc tta ttt gac ttt ctc ctc    336
Phe Leu Leu Gly Thr Trp Val Leu Ser Ala Leu Phe Asp Phe Leu Leu
65                  70                  75
att gaa gct atg cag tat ttc ttt ggc atc act gca gct agt aat ttg    384
Ile Glu Ala Met Gln Tyr Phe Phe Gly Ile Thr Ala Ala Ser Asn Leu
80                  85                  90                  95
cct tct gga tta atc ttt tgt tgt gct ttt tgc tct gag act aaa ctc    432
Pro Ser Gly Leu Ile Phe Cys Cys Ala Phe Cys Ser Glu Thr Lys Leu
                    100                 105                 110
ttc tta tca aga caa gct atg gca gag aac ttt tcc atc taataaattt      481
Phe Leu Ser Arg Gln Ala Met Ala Glu Asn Phe Ser Ile
                    115                 120
aagagtagat tcatctgtat ggttgagagt aggctctgac tatgtatatg tgtataataa   541
acctacatat ccaaaaaaaa aaaaaaa                                       568
```

```
<210> 41
<211> 569
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 168..332

<221> polyA_signal
<222> 557..562

<400> 41
```

```
aggggcgtg gggccatggt ggtcttgcgg gcggggaaga agacctttct ccccctctc    60
tgccgcgcct tcgcctgccg cggctgtcaa ctcgctccgg agcgcggcgc cgagcgcagg   120
gatacggcgc ccagcggggt cagaaagcaa cattgaatgc agaagaa atg gcg gac    176
                                                      Met Ala Asp
                                                      1
ttc tac aag gaa ttt tta agt aaa aat ttt cag aag cgc atg tat tat    224
Phe Tyr Lys Glu Phe Leu Ser Lys Asn Phe Gln Lys Arg Met Tyr Tyr
          5                   10                  15
aac aga gat tgg tac aag cgc aat ttt gcc atc acc ttc ttc atg gga    272
Asn Arg Asp Trp Tyr Lys Arg Asn Phe Ala Ile Thr Phe Phe Met Gly
20                  25                  30                  35
aaa gtg gcc ctg gaa agg att tgg aac aag ctt aaa cag aaa caa aag    320
Lys Val Ala Leu Glu Arg Ile Trp Asn Lys Leu Lys Gln Lys Gln Lys
                    40                  45                  50
aag agg agc aac taggagtcca ctctgaccca gccagagtcc aggtttccac         372
Lys Arg Ser Asn
                55
aggaagcaga tggagctcct ttcacagggg ctctgagaaa aactggagcc gatctcaaga   432
agccccacat cttcctaagg ggccccatgg cctgtttggg ggcagggtag gtcctggggc   492
actgtgggcc gcctgcctgc tgatgtgggc tctaggccag cttgttgtca cgtacgtggt   552
gtgaaataaa gcccaag                                                 569
```

```
<210> 42
<211> 895
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 51..251
```

<221> sig_peptide
<222> 51..110
<223> Von Heijne matrix
      score 5.3
      seq ALIFGGFISLIGA/AF

<221> polyA_signal
<222> 849..854

<221> polyA_site
<222> 882..895

<400> 42

```
ccgagagtgc cgggcggtcg gcgggtcagg gcagcccggg gcctgacgcc atg tcc        56
                                                         Met Ser
                                                         -20
cgg aac ctg cgc acc gcg ctc act ttc ggc ggc ttc atc tcc ctg atc      104
Arg Asn Leu Arg Thr Ala Leu Ile Phe Gly Gly Phe Ile Ser Leu Ile
        -15                 -10                 -5
ggc gcc gcc ttc tat ccc atc tac ttc cgg ccc cta atg aga ttg gag      152
Gly Ala Ala Phe Tyr Pro Ile Tyr Phe Arg Pro Leu Met Arg Leu Glu
          1               5                   10
gag tac aag aag gaa caa gct ata aat cgg gct gga att gtt caa gag      200
Glu Tyr Lys Lys Glu Gln Ala Ile Asn Arg Ala Gly Ile Val Gln Glu
15                  20                  25                  30
gat gtg cag cca cca ggg tta aaa gtg tgg tct gat cca ttt ggc agg      248
Asp Val Gln Pro Pro Gly Leu Lys Val Trp Ser Asp Pro Phe Gly Arg
                35                  40                  45
aaa tgagagggct gtcatcagct ctgattaaga aaggagattt cttcatgctt          301
Lys
tcgattctgc atggggtaca gccagtcacc tcaccagaga atgacggctg gagaagaaaa    361
ctctgtaata ccataaataa gagtgcttgt aataaaagac tgtgcacaag gattaatatt    421
tcccttctta agtatcaaaa gaactctgga acaaattata ccattaggaa ggttttcatg    481
attcagttga ttttccaaaa atgaagctat ctcacccagc tgggtttgga ggagcaatct    541
gcttattatt ctgtcgttac cacttactca agcgagctgt gatatgaata caagcaacca    601
gtgggctcgg gaaggtccgg gtctcttctg ccatcttcca gataagagat ttcagtaaaa    661
aactgccatg ctgagctgcc ttatagagct cttcgaaaat gttcgagttg ataaagctct    721
ttgaggacaa ggtacttcgt gcacctcatg ctgaagattg caccatgttg gaagataaat    781
atgaagcaag tcaaactaga tgcatacact tgtgtagaaa tcaataatca attaatagaa    841
gtgaaaaaat agacattaag atgatttatt ccactttgc aaaaaaaaaa aaaa          895
```

<210> 43
<211> 691
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 20..613

<221> sig_peptide
<222> 20..82
<223> Von Heijne matrix
      score 10
      seq LWALAMVTRPASA/AP

<400> 43

```
ataccttaga ccctcagtc atg cca gtg cct gct ctg tgc ctg ctc tgg gcc      52
                     Met Pro Val Pro Ala Leu Cys Leu Leu Trp Ala
```

```
                          -20                       -15
ctg gca atg gtg acc cgg cct gcc tca gcg gcc ccc atg ggc ggc cca        100
Leu Ala Met Val Thr Arg Pro Ala Ser Ala Ala Pro Met Gly Gly Pro
-10             -5                  1             5
gaa ctg gca cag cat gag gag ctg acc ctg ctc ttc cat ggg acc ctg        148
Glu Leu Ala Gln His Glu Glu Leu Thr Leu Leu Phe His Gly Thr Leu
            10                  15                  20
cag ctg ggc cag gcc ctc aac ggt gtg tac agg acc acg gag gga tgg        196
Gln Leu Gly Gln Ala Leu Asn Gly Val Tyr Arg Thr Thr Glu Gly Trp
        25                  30                  35
ctg aca aag gcc agg aac agc ctg ggt ctc tat ggc cgc aca ata gaa        244
Leu Thr Lys Ala Arg Asn Ser Leu Gly Leu Tyr Gly Arg Thr Ile Glu
    40                  45                  50
ctc ctg ggg cag gag gtc agc cgg ggc cgg gat gca gcc cag gaa ctt        292
Leu Leu Gly Gln Glu Val Ser Arg Gly Arg Asp Ala Ala Gln Glu Leu
55                  60                  65                  70
cgg gca agc ctg ttg gag act cag atg gag gag gat att ctg cag ctg        340
Arg Ala Ser Leu Leu Glu Thr Gln Met Glu Glu Asp Ile Leu Gln Leu
                75                  80                  85
cag gca gag gcc aca gct gag gtg ctg ggg gag gtg gcc cag gca cag        388
Gln Ala Glu Ala Thr Ala Glu Val Leu Gly Glu Val Ala Gln Ala Gln
                90                  95                  100
aag gtg cta cgg gac agc gtg cag cgg cta gaa gtc cag ctg agg agc        436
Lys Val Leu Arg Asp Ser Val Gln Arg Leu Glu Val Gln Leu Arg Ser
        105                 110                 115
gcc tgg ctg ggc cct gcc tac cga gaa ttt gag gtc tta aag gct cac        484
Ala Trp Leu Gly Pro Ala Tyr Arg Glu Phe Glu Val Leu Lys Ala His
    120                 125                 130
gct gac aag cag agc cac atc cta tgg gcc ctc aca ggc cac gtg cag        532
Ala Asp Lys Gln Ser His Ile Leu Trp Ala Leu Thr Gly His Val Gln
135                 140                 145                 150
cgg cag agg cgg gag atg gtg gca cag cag cat cgg ctg cga cag atc        580
Arg Gln Arg Arg Glu Met Val Ala Gln Gln His Arg Leu Arg Gln Ile
                155                 160                 165
cag gag aga ctc cac aca gcg gcg ctc cca gcc tgaatctgcc tggatggaac      633
Gln Glu Arg Leu His Thr Ala Ala Leu Pro Ala
        170                 175
tgaggaccaa tcatgctgca aggaacactt ccacgccccg tgaggcccct gtgcaggg       691
```

```
<210> 44
<211> 458
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 12..416

<221> sig_peptide
<222> 12..86
<223> Von Heijne matrix
      score 4
      seq LVVMVPLVGLIHL/GW

<221> polyA_signal
<222> 425..430

<221> polyA_site
<222> 445..458
```

```
<400> 44
gctgaagtac t atg agc ctt cgg aac ttg tgg aga gac tac aaa gtt ttg        50
           Met Ser Leu Arg Asn Leu Trp Arg Asp Tyr Lys Val Leu
                -25                 -20                 -15
gtt gtt atg gtc cct tta gtt ggg ctc ata cat ttg ggg tgg tac aga        98
Val Val Met Val Pro Leu Val Gly Leu Ile His Leu Gly Trp Tyr Arg
         -10                 -5                   1
atc aaa agc agc cct gtt ttc caa ata cct aaa aac gac gac att cct       146
Ile Lys Ser Ser Pro Val Phe Gln Ile Pro Lys Asn Asp Asp Ile Pro
5                   10                  15                  20
gag caa gat agt ctg gga ctt tca aat ctt cag aag agc caa atc cag       194
Glu Gln Asp Ser Leu Gly Leu Ser Asn Leu Gln Lys Ser Gln Ile Gln
                25                  30                  35
ggg aag nta gca ggc ttg caa tct tca ggt aaa gaa gca gct ttg aat       242
Gly Lys Xaa Ala Gly Leu Gln Ser Ser Gly Lys Glu Ala Ala Leu Asn
                40                  45                  50
ctg agc ttc ata tcg aaa gaa gag atg aaa aat acc agt tgg att aga       290
Leu Ser Phe Ile Ser Lys Glu Glu Met Lys Asn Thr Ser Trp Ile Arg
                55                  60                  65
aag aac tgg ctt ctt gta gct ggg ata tct ttc ata ggt gac cat ctt       338
Lys Asn Trp Leu Leu Val Ala Gly Ile Ser Phe Ile Gly Asp His Leu
        70                  75                  80
gga aca tac ttt ttg cag agg tct gca aag cag tct gta aaa ttt cag       386
Gly Thr Tyr Phe Leu Gln Arg Ser Ala Lys Gln Ser Val Lys Phe Gln
85                  90                  95                  100
tct caa agc aaa caa aag agt att gaa gag tgaagtaaaa taaatatttg         436
Ser Gln Ser Lys Gln Lys Ser Ile Glu Glu
                105                 110
gaattactaa aaaaaaaaaa aa                                               458
```

```
<210> 45
<211> 2036
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 276..1040

<221> sig_peptide
<222> 276..485
<223> Von Heijne matrix
      score 3.9
      seq SVIGVMLAPFTAG/LS

<221> polyA_site
<222> 2024..2036
```

```
<400> 45
gatcctgggt gcagctcatc acaagcgtcg gggtgcagca aaaccatcca ggctggacag      60
tggctggaca gttccaagaa aagaaacgct tcaccgaaga agtcattgaa tacttccaga     120
agaaagttag cccagtgcat ctgaaaatcc tgctgactag cgatgaagcc tggaagagat     180
tcgtgcgtgt ggctggattg cccagggaag aagcagatgc tctctatgaa gctctgaaga     240
atcttacacc atatgtggct attgaggaca aagac atg cag caa aaa gaa cag        293
                                       Met Gln Gln Lys Glu Gln
                                       -70                 -65
cag ttt agg gag tgg ttt ttg aaa gag ttt cct caa atc aga tgg aag       341
Gln Phe Arg Glu Trp Phe Leu Lys Glu Phe Pro Gln Ile Arg Trp Lys
        -60                 -55                 -50
att cag gag tcc ata gaa agg ctt cgt gtc att gca aat gag att gaa       389
```

```
Ile Gln Glu Ser Ile Glu Arg Leu Arg Val Ile Ala Asn Glu Ile Glu
         -45                 -40                 -35
aag gtc cac aga ggc tgc gtc atc gcc aat gtg gtg tct ggc tcc act    437
Lys Val His Arg Gly Cys Val Ile Ala Asn Val Val Ser Gly Ser Thr
         -30                 -25                 -20
ggc atc ctg tct gtc att ggc gtt atg ttg gca cca ttt aca gca ggg    485
Gly Ile Leu Ser Val Ile Gly Val Met Leu Ala Pro Phe Thr Ala Gly
         -15                 -10                  -5
ctg agc ctg agc att act gca gct ggg gta ggg ctg gga ata gca tct    533
Leu Ser Leu Ser Ile Thr Ala Ala Gly Val Gly Leu Gly Ile Ala Ser
1                 5                  10                  15
gcc acg gct ggg atc gcc tcc agc atc gtg gag aac aca tac aca agg    581
Ala Thr Ala Gly Ile Ala Ser Ser Ile Val Glu Asn Thr Tyr Thr Arg
             20                  25                  30
tca gca gaa ctc aca gcc agc agg ctg act gca acc agc act gac caa    629
Ser Ala Glu Leu Thr Ala Ser Arg Leu Thr Ala Thr Ser Thr Asp Gln
             35                  40                  45
ttg gag gca tta agg gac att ctg cat gac atc aca ccc aat gtg ctt    677
Leu Glu Ala Leu Arg Asp Ile Leu His Asp Ile Thr Pro Asn Val Leu
         50                  55                  60
tcc ttt gca ctt gat ttt gac gaa gcc aca aaa atg att gcg aat gat    725
Ser Phe Ala Leu Asp Phe Asp Glu Ala Thr Lys Met Ile Ala Asn Asp
65                  70                  75                  80
gtc cat aca ctc agg aga tct aaa gcc act gtt gga cgc cct ttg att    773
Val His Thr Leu Arg Arg Ser Lys Ala Thr Val Gly Arg Pro Leu Ile
                 85                  90                  95
gct tgg cga tat gta cct ata aat gtt gtt gag aca ctg aga aca cgt    821
Ala Trp Arg Tyr Val Pro Ile Asn Val Val Glu Thr Leu Arg Thr Arg
             100                 105                 110
ggg gcc ccc acc cgg ata gtg aga aaa gta gcc cgg aac ctg ggc aag    869
Gly Ala Pro Thr Arg Ile Val Arg Lys Val Ala Arg Asn Leu Gly Lys
             115                 120                 125
gcc act tca ggt gtc ctc gtt gtg ctg gat gta gtc aac ctt gtg caa    917
Ala Thr Ser Gly Val Leu Val Val Leu Asp Val Val Asn Leu Val Gln
             130                 135                 140
gac tca ctg gac ttg cac aag ggg gaa aaa tcc gag tct gct gag ttg    965
Asp Ser Leu Asp Leu His Lys Gly Glu Lys Ser Glu Ser Ala Glu Leu
145                 150                 155                 160
ctg agg cag tgg gct cag gag ctg gag gag aat ctc aat gag ctc acc   1013
Leu Arg Gln Trp Ala Gln Glu Leu Glu Glu Asn Leu Asn Glu Leu Thr
             165                 170                 175
cat atc cat cag agt cta aaa gca ggc taggcccaat tgttgcggga          1060
His Ile His Gln Ser Leu Lys Ala Gly
             180                 185
agtcagggac cccaaacgga gggactggct gaagccatgg cagaagaacg tggattgtga   1120
agatttcatg gacatttatt agttccccaa attaatactt ttataatttc ctatgcctgt   1180
ctttaccgca atctctaaac acaaattgtg aagatttcat ggacacttat cacttcccca   1240
atcaataccc ttgtgatttc ttatgcctgt ctttacttta atctcctaat cctgtcagct   1300
gaggagggtg tatgtcacct caggaccatg tgataattgc gttaactgca caaattgtag   1360
agcatgtgtg tttgaacaat atgaaatctg gcaccttga aaaaagaaca ggataacagc     1420
aatcgttcag gggataagag agataaacctt aaactctgac caacagtgag ccgggtggag   1480
cagagtcata tttcttttct ttcaaaagca aatgggagaa atatcgctga attctttttc   1540
tcagcaagga acatccctga gaaagagaat gcacccctga gggtgggtct ataaatggcc   1600
tccttgggtg tggccatctt ctatggtcga gactgtaggg atgaaataaa ccccagtctc   1660
ccatagtgct cccaggctta ttaggaagag gaaattcccg cctaataaat tttggtcaga   1720
ccggttgctc tcaaaaccct gtctcctgat aagatgttat caatgacaat ggtgcctgaa   1780
acctcattag caattttaat ttctccccgg tcctgtggtc ctgtgatctc accctgcctc   1840
cacttgcctt gtgatattct attaccttgt gaagtaggtg atctttgtga cccacaccct   1900
attcatacac tccctcccct tttggaagtc cctaataaaa acttgctggt tttgcagctt   1960
gtgaggcatc acggaaccta ctgatgtgtg atgtctcccc tggacaccta gctttaaaat   2020
ttcaaaaaaa aaaaaa                                                    2036
```

```
<210> 46
<211> 1276
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 443..619

<221> sig_peptide
<222> 443..589
<223> Von Heijne matrix
      score 7
      seq LICVVCLYIVCRC/GS

<221> polyA_site
<222> 1267..1276

<400> 46
gaggcactca cggcatttca ttgctacttt aattttcatt attatgggat tgattgctgt        60
cacagctact gctgcagtag ctggagttgc tttgcattcc acagtacaaa cagcagacta       120
tgtaaataat tggtagaaaa attctactct gctgtggaat taccaagata atatagacca       180
gaaactagct gatcaaatta atgatctcca caaaactgta atgtggctag gggatcatat       240
agttagttta gaatatagaa tgcggttaca atgtgattga aatacctctg atttttgcat       300
tactcctcat ctgtgtaatg aaacagagca tgagtgggaa aaagttaaga gatatttaaa       360
aggtcatact agaaatttat ctttggatat tgcaaagcta aaggaacaag tatttcaagc       420
ccctcagata catctgacac ta atg cca gga act gaa gtg ctt gaa gga gct       472
                          Met Pro Gly Thr Glu Val Leu Glu Gly Ala
                                        -45                     -40

aca gac gga tta gca gct att aac ctg cta aaa tgg atc aag aca ctt       520
Thr Asp Gly Leu Ala Ala Ile Asn Leu Leu Lys Trp Ile Lys Thr Leu
              -35                 -30                 -25

gga ggc tct gtg att tca atg att gtg ctt tta atc tgt gtt gtt tgt       568
Gly Gly Ser Val Ile Ser Met Ile Val Leu Leu Ile Cys Val Val Cys
          -20                 -15                 -10

ctt tat ata gtc tgt aga tgc gga agc cac ctc tgg aga gaa agc cac       616
Leu Tyr Ile Val Cys Arg Cys Gly Ser His Leu Trp Arg Glu Ser His
        -5                 1                 5

cac tgagagcaag caatgatagc tgtggcggtt ttgcaaaaag aaaagggaga           669
His
10
caagcgccca gctatagtta ccaataaagc atggtactgg tattaaaata ggcatgtgtt       729
ctgttccaat ggaacagaat agagaaccca gaaacaaagc caaatattta cagccaactg       789
atctctgaca aagcaaacaa aaacataaag tggggaaagg acaccctatt ccacaaatag       849
tgcagggata attggcaagc cacatgtaga aaaatgaagc tggatcctcg tctctcactt       909
tatacaaaaa tcaactcaaa atgggtcaaa gtcttaactc taagacctga aaccataaca       969
attctagaaa ataacattgg aaaaactctt ctagacattg gtttaggcaa aaagttcatg      1029
accaagaacc caaaagcaaa tgcaataaaa aggaagataa atagatggga cctaattaag      1089
ctgaaaagct tctgcatagc aaaaggaata atcagcagag caaacagaca acccacaggg      1149
tgggagaaaa tatttgcaag ctatgtatct gacaatggac taatatccag aatctacaag      1209
gaattcaaac aattagcaag aaaaaacact tgtattgtgt ttgctctgta aatcagcaaa      1269
aaaaaaa                                                                1276


<210> 47
<211> 747
<212> DNA
<213> Homo sapiens
```

<220>
<221> CDS
<222> 206..745

<400> 47
accagaagca ggtgatttcc gagctcagca atgctcagct cataatgatg tcaagcacca   60
tggccagttt tatgaatggc ttcctgtgtc taatgaccct gacaacccat gttcactcaa   120
gtgccaagcc aaaggaacaa ccctggttgt tgaactagca cctaaggtct tagatggtac   180
gcgttgctat acagaatctt tggat atg tgc atc agt ggt tta tgc caa att   232
                            Met Cys Ile Ser Gly Leu Cys Gln Ile
                            1   5

gtt ggc tgc gat cac cag ctg gga agc acc gtc aag gaa gat aac tgt   280
Val Gly Cys Asp His Gln Leu Gly Ser Thr Val Lys Glu Asp Asn Cys
10   15   20   25

ggg gtc tgc aac gga gat ggg tcc acc tgc cgg ctg gtc cga ggg cag   328
Gly Val Cys Asn Gly Asp Gly Ser Thr Cys Arg Leu Val Arg Gly Gln
                30   35   40

tat aaa tcc cag ctc tcc gca acc aaa tcg gat gat act gtg gtt gca   376
Tyr Lys Ser Gln Leu Ser Ala Thr Lys Ser Asp Asp Thr Val Val Ala
            45   50   55

att ccc tat gga agt aga cat att cgc ctt gtc tta aaa ggt cct gat   424
Ile Pro Tyr Gly Ser Arg His Ile Arg Leu Val Leu Lys Gly Pro Asp
        60   65   70

cac tta tat ctg gaa acc aaa acc ctc cag ggg act aaa ggt gaa aac   472
His Leu Tyr Leu Glu Thr Lys Thr Leu Gln Gly Thr Lys Gly Glu Asn
    75   80   85

agt ctc agc tcc aca gga act ttc ctt gtg gac aat tct agt gtg gac   520
Ser Leu Ser Ser Thr Gly Thr Phe Leu Val Asp Asn Ser Ser Val Asp
90   95   100   105

ttc cag aaa ttt cca gac aaa gag ata ctg aga atg gct gga cca ctc   568
Phe Gln Lys Phe Pro Asp Lys Glu Ile Leu Arg Met Ala Gly Pro Leu
                110   115   120

aca gca gat ttc att gtc aag att cgt aac tcg ggc tcc gct gac agt   616
Thr Ala Asp Phe Ile Val Lys Ile Arg Asn Ser Gly Ser Ala Asp Ser
            125   130   135

aca gtc cag ttc atc ttc tat caa ccc atc atc cac cga tgg agg gag   664
Thr Val Gln Phe Ile Phe Tyr Gln Pro Ile Ile His Arg Trp Arg Glu
        140   145   150

acg gat ttc ttt cct tgc tca gca acc tgt gga gga ggt tat cag ctg   712
Thr Asp Phe Phe Pro Cys Ser Ala Thr Cys Gly Gly Gly Tyr Gln Leu
    155   160   165

aca tcg gct gag tgc tac gat ctg agg agc aac cg   747
Thr Ser Ala Glu Cys Tyr Asp Leu Arg Ser Asn
170   175   180

<210> 48
<211> 561
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 36..521

<221> sig_peptide
<222> 36..104
<223> Von Heijne matrix
      score 7.4
      seq VLLLAALPPVLLP/GA

<221> polyA_signal
<222> 528..533

<221> polyA_site
<222> 548..561

<400> 48
gacgcctctt tcagcccggg atcgccccag caggg atg ggc gac aag atc tgg   53
                                        Met Gly Asp Lys Ile Trp
                                                              -20

ctg ccc ttc ccc gtg ctc ctt ctg gcc gct ctg cct ccg gtg ctg ctg   101
Leu Pro Phe Pro Val Leu Leu Leu Ala Ala Leu Pro Pro Val Leu Leu
        -15   -10   -5

cct ggg gcg gcc ggc ttc aca cct tcc ctc gat agc gac ttc acc ttt   149
Pro Gly Ala Ala Gly Phe Thr Pro Ser Leu Asp Ser Asp Phe Thr Phe
1   5   10   15

acc ctt ccc gcc ggc cag aag gag tgc ttc tac cag ccc atg ccc ctg   197
Thr Leu Pro Ala Gly Gln Lys Glu Cys Phe Tyr Gln Pro Met Pro Leu
                20   25   30

aag gcc tcg ctg gag atc gag tac caa gtt tta gat gga gca gga tta   245
Lys Ala Ser Leu Glu Ile Glu Tyr Gln Val Leu Asp Gly Ala Gly Leu
            35   40   45

gat att gat ttc cat ctt gcc tct cca gaa ggc aaa acc tta gtt ttt   293
Asp Ile Asp Phe His Leu Ala Ser Pro Glu Gly Lys Thr Leu Val Phe
            50   55   60

gaa caa aga aaa tca gat gga gtt cac act gta gag act gaa gtt ggt   341
Glu Gln Arg Lys Ser Asp Gly Val His Thr Val Glu Thr Glu Val Gly
65   70   75

gat tac atg ttc tgc ttt gac aat aca ttc agc acc att tct gag aag   389
Asp Tyr Met Phe Cys Phe Asp Asn Thr Phe Ser Thr Ile Ser Glu Lys
80   85   90   95

gtg att ttc ttt gaa tta atc ccg gat aat atg gga gaa cag gca caa   437
Val Ile Phe Phe Glu Leu Ile Pro Asp Asn Met Gly Glu Gln Ala Gln
                100   105   110

gaa caa gaa gat tgg aag aaa tat att act ggc aca gat ata ttg gat   485
Glu Gln Glu Asp Trp Lys Lys Tyr Ile Thr Gly Thr Asp Ile Leu Asp
                115   120   125

atg aaa ctg gaa gac atc ctg gtc agt atg gtc ttc taataaaata   531
Met Lys Leu Glu Asp Ile Leu Val Ser Met Val Phe
            130   135

aaaattatta acagccaaaa aaaaaaaaaa   561

<210> 49
<211> 632
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 36..395

<221> sig_peptide
<222> 36..104
<223> Von Heijne matrix
      score 7.4
      seq VLLLAALPPVLLP/GA

<221> polyA_signal
<222> 599..604

```
<221> polyA_site
<222> 619..632

<400> 49
gacgcctctt tcagcccggg atcgccccag caggg atg ggc gac aag atc tgg      53
                                        Met Gly Asp Lys Ile Trp
                                                            -20
ctg ccc ttc ccc gtg ctc ctt ctg gcc gct ctg cct ccg gtg ctg ctg    101
Leu Pro Phe Pro Val Leu Leu Leu Ala Ala Leu Pro Pro Val Leu Leu
        -15             -10             -5
cct ggg gcg gcc ggc ttc aca cct tcc ctc gat agc gac ttc acc ttt    149
Pro Gly Ala Ala Gly Phe Thr Pro Ser Leu Asp Ser Asp Phe Thr Phe
 1               5               10              15
acc ctt ccc gcc ggc cag aag gag tgc ttc tac cag ccc atg ccc ctg    197
Thr Leu Pro Ala Gly Gln Lys Glu Cys Phe Tyr Gln Pro Met Pro Leu
         20              25              30
aag gcc tcg ctg gag atc gag tac caa gtt tta gat gga gca gga tta    245
Lys Ala Ser Leu Glu Ile Glu Tyr Gln Val Leu Asp Gly Ala Gly Leu
         35              40              45
gat att gat ttc cat ctt gcc tct cca gaa ggc aaa acc tta gtt ttt    293
Asp Ile Asp Phe His Leu Ala Ser Pro Glu Gly Lys Thr Leu Val Phe
         50              55              60
gaa caa aga aaa tca gat gga gtt cac acg tgt ata aga agt aaa aat    341
Glu Gln Arg Lys Ser Asp Gly Val His Thr Cys Ile Arg Ser Lys Asn
         65              70              75
ggg cca ggc act gcg gtt cac gcc tat aat ccc agc act ttc cga ggc    389
Gly Pro Gly Thr Ala Val His Ala Tyr Asn Pro Ser Thr Phe Arg Gly
80              85              90              95
caa gtg tagagactga agttggtgat tacatgttct gctttgacaa tacattcagc    445
Gln Val
accattctg agaaggtgat tttctttgaa ttaatcctgg ataatatggg agaacaggca   505
caaggacaag aagattggaa gaaatatatt actggcacag atatattgga tatgaaactg   565
gaagacatcc tggtcagtat ggtcttctaa taaaataaaa attattaaca gccaaaaaaa   625
aaaaaaa                                                             632



<210> 50
<211> 370
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 21..41

<221> polyA_signal
<222> 328..333

<221> polyA_site
<222> 357..370

<400> 50
ctgggacttc tggcctcaca atg gtt gag atg act ggg gtg tagcagtgcc        51
                        Met Val Glu Met Thr Gly Val
                        1               5
aagtcgaggc tgtgaaaggc cttccacctt tactctcgtg ctcgtgccct ccccccattgt  111
taggagaagg gcatgctcag gccagcccat tagcccagga ggaggacaag aaacacacgg   171
agcagacaca agccacctca ccaacccagc caaggctgtc ctgaattagc aaccctgaca   231
cgtgtgagca agtccaacgg acaccggaag atccacctag tcaagcccaa ccaagactgg   291
cagagctgcc aagctgacca cttaaggcgc atgaggaata aacactcgtt gctgcatgcc   351
attgcaaaaa aaaaaaaaaa                                               370
```

<210> 51
<211> 994
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 35..631

<221> sig_peptide
<222> 35..160
<223> Von Heijne matrix
      score 8.6
      seq ASLFLLLSLTVFS/IV

<221> polyA_signal
<222> 901..906

<221> polyA_site
<222> 979..994

<400> 51
ataattggag ctgcaaagca gatcgtgaca agag atg gac ggt cag aag aaa aat   55
                                   Met Asp Gly Gln Lys Lys Asn
                                   -40

tgg aag gac aag gtt gtt gac ctc ctg tac tgg aga gac att aag aag   103
Trp Lys Asp Lys Val Val Asp Leu Leu Tyr Trp Arg Asp Ile Lys Lys
-35   -30   -25   -20

act gga gtg gtg ttt ggt gcc agc cta ttc ctg ctg ctt tca ttg aca   151
Thr Gly Val Val Phe Gly Ala Ser Leu Phe Leu Leu Leu Ser Leu Thr
                -15   -10   -5

gta ttc agc att gtg agc gta aca gcc tac att gcc ttg gcc ctg ctc   199
Val Phe Ser Ile Val Ser Val Thr Ala Tyr Ile Ala Leu Ala Leu Leu
            1   5   10

tct gtg acc atc agc ttt agg ata tac aag ggt gtg atc caa gct atc   247
Ser Val Thr Ile Ser Phe Arg Ile Tyr Lys Gly Val Ile Gln Ala Ile
    15   20   25

cag aaa tca gat gaa ggc cac cca ttc agg gca tat ctg gaa tct gaa   295
Gln Lys Ser Asp Glu Gly His Pro Phe Arg Ala Tyr Leu Glu Ser Glu
30   35   40   45

gtt gct ata tct gag gag ttg gtt cag aag tac agt aat tct gct ctt   343
Val Ala Ile Ser Glu Glu Leu Val Gln Lys Tyr Ser Asn Ser Ala Leu
            50   55   60

ggt cat gtg aac tgc acg ata aag gaa ctc agg cgc ctc ttc tta gtt   391
Gly His Val Asn Cys Thr Ile Lys Glu Leu Arg Arg Leu Phe Leu Val
            65   70   75

gat gat tta gtt gat tct ctg aag ttt gca gtg ttg atg tgg gta ttt   439
Asp Asp Leu Val Asp Ser Leu Lys Phe Ala Val Leu Met Trp Val Phe
            80   85   90

acc tat gtt ggt gcc ttg ttt aat ggt ctg aca cta ctg att ttg gct   487
Thr Tyr Val Gly Ala Leu Phe Asn Gly Leu Thr Leu Leu Ile Leu Ala
    95   100   105

ctc att tca ctc ttc agt gtt cct gtt att tat gaa cgg cat cag gca   535
Leu Ile Ser Leu Phe Ser Val Pro Val Ile Tyr Glu Arg His Gln Ala
110   115   120   125

cag ata gat cat tat cta gta ctt gca aat aag aat gtt aaa gat gct   583
Gln Ile Asp His Tyr Leu Val Leu Ala Asn Lys Asn Val Lys Asp Ala
            130   135   140

atg gct aaa atc caa gca aaa atc cct gga ttg aag cgc aaa gct gaa   631

```
Met Ala Lys Ile Gln Ala Lys Ile Pro Gly Leu Lys Arg Lys Ala Glu
         145                 150                 155
tgaaaacgcc caaaataatt agtaggagtt catctttaaa ggggatattc atttgattat      691
acggggggagg gtcagggaag aacgaacctt gacgttgcag tgcagtttca cagatcgttg      751
ttagatcttt attttttagcc atgcactgtt gtgaggaaaa attacctgtc ttgactgcca      811
tgtgttcatc atcttaagta ttgtaagctg ctatgtatgg atttaaaccg taatcatatc      871
tttttcctat ctatctgagg cactggtgga ataaaaaacc tgtatatttt actttgttgc      931
agatagtctt gccgcatctt ggcaagttgc agagatggtg gagctagaaa aaaaaaaaac      991
aaa                                                                    994
```


```
<210> 52
<211> 412
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 271..399
```

```
<400> 52
gccgctagcg cctcgagcga tgcacctcct ttccaactgg caaaccccg cttccagcag      60
acgtccttct atggccgctt caggcacttc ttggatatca tcgaccctcg cacactcttt     120
gtcactgaga gacgtctcag agaggctgtg cagctgctgg aggactataa gcatgggacc     180
ctgcgcccgg gggtcaccaa tgaacagctc tggagtgcac agaaaatcaa gcaggctatt     240
ctacatccgg acaccaatga gaagatcttc atg cca ttt aga atg tca ggt tat     294
                                    Met Pro Phe Arg Met Ser Gly Tyr
                                     1                   5
att cct ttt ggg acg cca att gta agt gtt acc ttc aaa gga ttt cct     342
Ile Pro Phe Gly Thr Pro Ile Val Ser Val Thr Phe Lys Gly Phe Pro
    10                  15                  20
ttt cta aaa aat tat ttt aaa tgt cta act tta tgt tat tgc tca cgg     390
Phe Leu Lys Asn Tyr Phe Lys Cys Leu Thr Leu Cys Tyr Cys Ser Arg
25                  30                  35                  40
gta ttt gac tgaattgttg att                                            412
Val Phe Asp
```


```
<210> 53
<211> 597
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 103..252

<221> sig_peptide
<222> 103..213
<223> Von Heijne matrix
      score 3.9
      seq PGPSLRLFSGSQA/SV

<221> polyA_site
<222> 588..597
```

```
<400> 53
gaaaggtcag aggaaggagc tgtgggaagc tcgcagcagg tatcggagct taagccagtg      60
gatttggggg ccctgggctc cctagccggc tgcggtgtga ga atg gag tgg gca      114
                                                Met Glu Trp Ala
```

```
                                                                            -35
gga aag cag cgg gac ttt cag gta agg gca gct ccg ggc tgg gat cat         162
Gly Lys Gln Arg Asp Phe Gln Val Arg Ala Ala Pro Gly Trp Asp His
          -30                       -25                       -20
ttg gcc tcc ttt cct ggc cct tct ctc cgg ctg ttt tct ggg agt cag         210
Leu Ala Ser Phe Pro Gly Pro Ser Leu Arg Leu Phe Ser Gly Ser Gln
          -15                       -10                       -5
gcg agt gtc tgt agt ctc tgc tcg ggg ttt ggg gct cag gaa             252
Ala Ser Val Cys Ser Leu Cys Ser Gly Phe Gly Ala Gln Glu
  1               5                   10
tgatgtcatg ctccaacagt tggattctat tagcttaagg aggagggaaa cagccaattt        312
tcttgacttt gcaaatctag ctgatctcac tcttgctgaa tctgaggtgt ttagacttca        372
ctctaaaaag catcatttta cttttattta gcacaaaggc acaggatatt tttacaggaa        432
gaatctttta tatggaaaaa tctgagttaa catcactccc gtggtgtttg tagttcttac        492
agggaaactc cagtgccttt tgagccgctt gttcgtccta gtgaacactg tctgttttgt        552
ctcttggtgc tgctatgtct gacctgtaat gggagaaaaa aagaa                        597
```

<210> 54
<211> 748
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 2..460

<221> polyA_signal
<222> 713..718

<221> polyA_site
<222> 735..748

<400> 54
```
c aca gtt cct ctc ctc cta gag cct gcc gac cat gcc cgc ggg cgt gcc      49
  Thr Val Pro Leu Leu Leu Glu Pro Ala Asp His Ala Arg Gly Arg Ala
  1           5                   10                      15
cat gtc cac cta cct gaa aat gtt cgc agc cag tct cct ggc cat gtg        97
His Val His Leu Pro Glu Asn Val Arg Ser Gln Ser Pro Gly His Val
              20                      25                      30
cgc agg ggc aga agt ggt gca cag gta cta ccg acc gga cct gat gag        145
Arg Arg Gly Arg Ser Gly Ala Gln Val Leu Pro Thr Gly Pro Asp Glu
              35                      40                      45
aaa cag gtt gag aag agt gaa gtt gat ttc tca aag tca cat agc tta        193
Lys Gln Val Glu Lys Ser Glu Val Asp Phe Ser Lys Ser His Ser Leu
  50                      55                      60
gtg aga cga ttt gag gat ctg aag ccc aag ctt tct gtt tgc aaa act        241
Val Arg Arg Phe Glu Asp Leu Lys Pro Lys Leu Ser Val Cys Lys Thr
  65                      70                      75                  80
gga tca caa gtc ttt cgg tcg gag aac tgg aag gtc tgg gca gag tcg        289
Gly Ser Gln Val Phe Arg Ser Glu Asn Trp Lys Val Trp Ala Glu Ser
              85                      90                      95
agc aga gga gac cat gat gac tgc cta gac ttg tgc tca gtg ctg tgt        337
Ser Arg Gly Asp His Asp Asp Cys Leu Asp Leu Cys Ser Val Leu Cys
              100                     105                     110
tgg gga gaa ctg cta cgg aca ata cct gaa att cca cca aag cgt gga        385
Trp Gly Glu Leu Leu Arg Thr Ile Pro Glu Ile Pro Pro Lys Arg Gly
              115                     120                     125
gaa ctc aaa acg gag ctt ttg gga ctg aaa gaa aga aaa cac aaa cct        433
Glu Leu Lys Thr Glu Leu Leu Gly Leu Lys Glu Arg Lys His Lys Pro
              130                     135                     140
```

```
caa gtt tct caa cag gag gaa ctt aaa taactatgcc aagaattctg        480
Gln Val Ser Gln Gln Glu Glu Leu Lys
145               150
tgaataatat aagtcttaaa tatgtatttc ttaatttatt gcatcaaact acttgtcctt  540
aagcacttag tctaatgcta actgcaagag gaggtgctca gtggatgttt agccgatacg  600
ttgaaattta attacggttt gattgatatt tcttgaaaac cgccaaagca catatcatca  660
aaccatttca tgaatatggt ttggaagatg tttagtcttg aatataatgc gaaatagaat  720
atttgtaagt ctaccaaaaa aaaaaaaa                                     748
```


```
<210> 55
<211> 703
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 31..231

<221> polyA_signal
<222> 769..774

<221> polyA_site
<222> 690..703

<400> 55
```

```
ctctggtggc tctgctacgg cggcgcagaa atg agg cag aag cgg aaa gga gat   54
                                 Met Arg Gln Lys Arg Lys Gly Asp
                                 1               5
ctc agc cct gct aag ctg atg atg ctg act ata gga gat gtt att aaa   102
Leu Ser Pro Ala Lys Leu Met Met Leu Thr Ile Gly Asp Val Ile Lys
    10              15                  20
caa ctg att gaa gcc cac gag cag ggg aaa gac atc gat cta aat aag   150
Gln Leu Ile Glu Ala His Glu Gln Gly Lys Asp Ile Asp Leu Asn Lys
25              30                  35                  40
gtg aga acc aag aca gct gcc aaa tat ggc ctt tct gcc cag ccc cgc   198
Val Arg Thr Lys Thr Ala Ala Lys Tyr Gly Leu Ser Ala Gln Pro Arg
                45                  50                  55
ctg gtg gat atc att gct gcc gtc cct cct gag tagctgggat tacaggcacc  251
Leu Val Asp Ile Ile Ala Ala Val Pro Pro Glu
                60                  65
cgccgctgcc aattttgta tttttagtag ggatgggggt ttcaccatat tggtcaggct  311
ggtctcgaac tcctgacctc aggtgatcaa cccaccttgg cctccctaaa tgccgggatt  371
acaggcatga gccaccgctc cgggcctttg attttttaag gtggattttg gttgttataa  431
atggagaaag gtaagagttc aagttcaacc cgtgtgtgaa agcaaaacaa tggaaaacag  491
gattggcttc ttcaaaggct cctcttgtag aactgcctct ttgaaatttc gaggtaatct  551
actttggaga ctctgcctgg agaggtcag ttcctaagtt aaaagcatcg cttaaccttg   611
gctcctgtgg cattttacaa aggtttaaag gaattgattc ctctgaaagg gcctgaaaat  671
aaaaagtctt taacatacaa aaaaaaaaa aa                                 703
```


```
<210> 56
<211> 725
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 305..565
```

```
<221> polyA_signal
<222> 694..699

<221> polyA_site
<222> 713..725

<400> 56
ctcacggtgg tgaaggtcac agggttgcag cactcccagt agaccaggag ctccgggagg     60
cagggccggc cccacgtcct ctgcgcacca ccctgagttg gatcctctgt gcgccacccc    120
tgagttggat ccagggctag ctgctgttga cctccccact cccacgctgc cctcctgcct    180
gcagccatga cgcccctgct caccctgatc ctggtggtcc tcatgggctt acctctggcc    240
caggccttgg actgccacgt gtgaggacta caaatccctc caggatatca ttgccatcct    300
gggt atg gat gaa ctt tct gag gaa gac aag ttg acc gtg tcc cgt gca    349
     Met Asp Glu Leu Ser Glu Glu Asp Lys Leu Thr Val Ser Arg Ala
      1           5                  10                  15
cgg aaa ata cag cgt ttc ttg tct cag cca ttc cag gtt gct gag gtc    397
Arg Lys Ile Gln Arg Phe Leu Ser Gln Pro Phe Gln Val Ala Glu Val
                20                  25                  30
ttc aca ggt cat atg ggg aag ctg gta ccc ctg aag gag acc atc aaa    445
Phe Thr Gly His Met Gly Lys Leu Val Pro Leu Lys Glu Thr Ile Lys
            35                  40                  45
gga ttc cag cag att ttg gca ggt gaa tat gac cat ctc cca gaa cag    493
Gly Phe Gln Gln Ile Leu Ala Gly Glu Tyr Asp His Leu Pro Glu Gln
        50                  55                  60
gcc ttc tat atg gtg gga ccc att gaa gaa gct gtg gca aaa gct gat    541
Ala Phe Tyr Met Val Gly Pro Ile Glu Glu Ala Val Ala Lys Ala Asp
65                  70                  75
aag ctg gct gaa gag cat tca tcg tgaggggtct ttgtcctctg tactgtctct    595
Lys Leu Ala Glu Glu His Ser Ser
80                  85
ctccttgccc ctaacccaaa aagcttcatt tttctgtgta ggctgcacaa gagccttgat    655
tgaagatata ttctttctga acagtattta aggtttccaa taaagtgtac acccctcaaa    715
aaaaaaaaaa                                                            725
```

```
<210> 57
<211> 1705
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 124..873

<221> sig_peptide
<222> 124..378
<223> Von Heijne matrix
      score 3.6
      seq HLSVVTLAAKVKC/IP

<221> polyA_signal
<222> 1673..1678

<221> polyA_site
<222> 1694..1705

<400> 57
cggaggtgag gagcggcggc cccgcccggt gcgctggagg tcgaagcttc caggtagcgg     60
cccgcagagc ctgacccagg ctctggacat cctgagccca agtcccccac actcagtgca    120
gtg atg agt gcg gaa gtg aag gtg aca ggg cag aac cag gag caa ttt    168
    Met Ser Ala Glu Val Lys Val Thr Gly Gln Asn Gln Glu Gln Phe
```

```
        -85                      -80                      -75
ctg ctc cta gcc aag tcg gcc aag ggg gca gcg ctg gcc aca ctc atc        216
Leu Leu Leu Ala Lys Ser Ala Lys Gly Ala Ala Leu Ala Thr Leu Ile
-70                      -65                      -60                      -55
cat cag gtg ctg gag gcc cct ggt gtc tac gtg ttt gga gaa ctg ctg        264
His Gln Val Leu Glu Ala Pro Gly Val Tyr Val Phe Gly Glu Leu Leu
                    -50                      -45                      -40
gac atg ccc aat gtt aga gag ctg naa gcc cgg aat ctt cct cca cta        312
Asp Met Pro Asn Val Arg Glu Leu Xaa Ala Arg Asn Leu Pro Pro Leu
                -35                      -30                      -25
aca gag gct cag aag aat aag ctt cga cac ctc tca gtt gtc acc ctg        360
Thr Glu Ala Gln Lys Asn Lys Leu Arg His Leu Ser Val Val Thr Leu
            -20                      -15                      -10
gct gct aaa gta aag tgt atc cca tat gca gtg ttg ctg gag gct ctt        408
Ala Ala Lys Val Lys Cys Ile Pro Tyr Ala Val Leu Leu Glu Ala Leu
        -5                       1                        5                       10
gcc ctg cgt aat gtg cgg cag ctg gaa gac ctt gtg att gag gct gtg        456
Ala Leu Arg Asn Val Arg Gln Leu Glu Asp Leu Val Ile Glu Ala Val
                15                       20                       25
tat gct gac gtg ctt cgt ggc tcc ctg gac cag cgc aac cag cgg ctc        504
Tyr Ala Asp Val Leu Arg Gly Ser Leu Asp Gln Arg Asn Gln Arg Leu
            30                       35                       40
gag gtt gac tac agc atc ggg cgg gac atc cag cgc cag gac ctc agt        552
Glu Val Asp Tyr Ser Ile Gly Arg Asp Ile Gln Arg Gln Asp Leu Ser
        45                       50                       55
gcc att gcc cga acc ctg cag gaa tgg tgt gtg ggc tgt gag gtc gtg        600
Ala Ile Ala Arg Thr Leu Gln Glu Trp Cys Val Gly Cys Glu Val Val
    60                       65                       70
ctg tca ggc att gag gag cag gtg agc cgt gcc aac caa cac aag gag        648
Leu Ser Gly Ile Glu Glu Gln Val Ser Arg Ala Asn Gln His Lys Glu
75                       80                       85                       90
cag cag ctg ggc ctg aag cag cag att gag agt gag gtt gcc aac ctt        696
Gln Gln Leu Gly Leu Lys Gln Gln Ile Glu Ser Glu Val Ala Asn Leu
                95                       100                      105
aaa aaa acc att aaa gtt acg acg gca gca gca gcc gca gcc aca tct        744
Lys Lys Thr Ile Lys Val Thr Thr Ala Ala Ala Ala Ala Ala Thr Ser
            110                      115                      120
cag gac cct gag caa cac ctg act gag ctg agg gaa cca gct cct ggc        792
Gln Asp Pro Glu Gln His Leu Thr Glu Leu Arg Glu Pro Ala Pro Gly
        125                      130                      135
acc aac cag cgc cag ccc agc aag aaa gcc tca aag ggc aag ggg ctc        840
Thr Asn Gln Arg Gln Pro Ser Lys Lys Ala Ser Lys Gly Lys Gly Leu
    140                      145                      150
cga ggg agc gcc aag att tgg tcc aag tcg aat tgaaagaact gtcgtttcct        893
Arg Gly Ser Ala Lys Ile Trp Ser Lys Ser Asn
155                      160                      165
ccctggggat gtggggtccc agctgcctgc ctgcctctta ggagtcctca gagagccttc       953
tgtgcccctg gccagctgat aatcctaggt tcatgaccct tcacctcccc taaccccaaa       1013
catagatcac accttctcta gggaggagtc aaatgtaggt catgtttttg ttggtacttt       1073
ctgttttttg tgacttcatg tgttccattg ctccccgctg ccatgctctc tcccttgttt       1133
ccttaagagc tcagcatctg tccctgttca ttacatgtca ttgagtaggt gggtagccct       1193
gatgggggtc gctctgtctg gagcataacc cacaggcgtt ttttctgcca ccccatccct       1253
gcatgcctga tccccagttc ctatacccta ccctgacct attgagcagc ctctgaagag        1313
ccatagggcc cccaccttta ctcacaccct gagaattctg ggagccagtc tgccatgcca       1373
ggagtcactg gacatgttca tcctagaatc ctgtcacact acagtcattt cttttcctct       1433
ctctggccct tgggtcctgg gaatgctgct gcttcaaccc cagagcctaa gaatggcagc       1493
cgtttcttaa catgttgaga gatgattctt tcttggccct ggccatctcg ggaagcttga       1553
tggcaatcct ggaagggttt aatctccttt tgtgagtttg gtggggaagg aagggtata        1613
tagattatat taaaaaaaaa aaggtatata tgcatatatc tatatataat atgacgcaga       1673
aataaatcta tgagaaatcc aaaaaaaaaa aa                                     1705
```

```
<210> 58
<211> 1069
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 135..206

<221> polyA_signal
<222> 850..855

<221> polyA_site
<222> 1056..1069

<400> 58
cccactccgc tctcacgact aagctctcac gattaaggca cgcctgcctc gattgtccag      60
cctctgccag aagaaagctt agcagccagc gcctcagtag agacctaagg gcgctgaatg     120
agtgggaaag ggaa atg ccg acc aat tgc gct gcg gcg ggc tgt gcc act       170
              Met Pro Thr Asn Cys Ala Ala Ala Gly Cys Ala Thr
               1               5                  10
acc tac aac aag cac att aac atc agc ttc cac agg taacctgggc            216
Thr Tyr Asn Lys His Ile Asn Ile Ser Phe His Arg
         15                  20
agggagtggg ggtgacggaa actggagttc ctattgtggc tatcgcttgt gtggaaggaa     276
caggaggatt ctgctaattc taataacttt cccagctggt agcagggaag catcgtatgt     336
cctttgtgtt tctcaaatct gcccaattgt tctctgcttt cggggaagct ttactcattt     396
tctaaaagaa atccaagtac tgtttggtca ttacccctta gtaaaaaaaa gtaacaggag     456
gatatcgtaa ttttctactg ttttattcct ctgttagacc gggccttgac atgaatgacg     516
ccgtaaggga gaaagagatc ttcccaatca gcaatcaccg taaaagcctg ctgtgttccc     576
gttaaaatta ggaaattctc actagatgaa ttgacatggg aggcatttag atttctaata     636
gtcacatagt aattctgcgg aggaattgag tcatctttga tagccatgga attaagcgat     696
gttaattaaa gtgcaaacga taacccttct gttcttacta gaatagagta ataaaaagaa     756
cctaggtttt cttttgtttg ctggaagaaa aatcaaaatt ctttagttct gtcaaaccag     816
aactcttgaa agcactttga acaatgcctg gaaaataaca ggtactctgt aaatgtttac     876
cttctctgca agtgcctgcc acgtgcccga agaaaagaca cattaaaaag ttaagtgaca     936
ccagtcctga ttttatatat tttatatacc taacaacgta tatgttagta tgtagaaatt     996
atatccttga ccttttttccc tacctattac gaactgtact tttattaaaa gctgccacta    1056
aaaaaaaaaa aaa                                                        1069


<210> 59
<211> 1084
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 135..818

<221> polyA_signal
<222> 909..914

<221> polyA_site
<222> 1071..1084

<400> 59
cccactccgc tctcacgact aagctctcac gattaaggca cgcctgcctc gattgtccag      60
cctctgccag aagaaagctt agcagccagc gcctcagtag aggcctaagg gcgctgaatg     120
agtgggaaag ggaa atg ccg acc aat tgc gct gcg gcg ggc tgt gcc act       170
```

```
                         Met Pro Thr Asn Cys Ala Ala Ala Gly Cys Ala Thr
                         1           5               10
acc tac aac aag cac att aac atc agc ttc cac agg ttt cct ttg gat        218
Thr Tyr Asn Lys His Ile Asn Ile Ser Phe His Arg Phe Pro Leu Asp
        15              20                  25
cct aaa aga aga aaa gaa tgg gtt cgc ctg gtt agg cgc aaa aat ttt        266
Pro Lys Arg Arg Lys Glu Trp Val Arg Leu Val Arg Arg Lys Asn Phe
    30              35                  40
gtg cca gga aaa cac act ttt ctt tgt tca aag cac ttt gaa gcc tcc        314
Val Pro Gly Lys His Thr Phe Leu Cys Ser Lys His Phe Glu Ala Ser
45                  50                  55                  60
tgt ttt gac cta aca gga caa act cga cga ctt aaa atg gat gct gtt        362
Cys Phe Asp Leu Thr Gly Gln Thr Arg Arg Leu Lys Met Asp Ala Val
                65                  70                  75
cca acc att ttt gat ttt tgt acc cat ata aag tct atg aaa ctc aag        410
Pro Thr Ile Phe Asp Phe Cys Thr His Ile Lys Ser Met Lys Leu Lys
                80                  85                  90
tca agg aat ctt ttg aag aaa aac aac agt tgt tct cca gct gga cca        458
Ser Arg Asn Leu Leu Lys Lys Asn Asn Ser Cys Ser Pro Ala Gly Pro
        95                  100                 105
tct agt tta aaa tca aac att agt agt cag caa gta cta ctt gaa cac        506
Ser Ser Leu Lys Ser Asn Ile Ser Ser Gln Gln Val Leu Leu Glu His
    110                 115                 120
agc tat gcc ttt agg aat cct atg gag gca aaa aag agg atc att aaa        554
Ser Tyr Ala Phe Arg Asn Pro Met Glu Ala Lys Lys Arg Ile Ile Lys
125                 130                 135                 140
ctg gaa aaa gaa ata gca agc tta aga aga aaa atg aaa act tgc cta        602
Leu Glu Lys Glu Ile Ala Ser Leu Arg Arg Lys Met Lys Thr Cys Leu
                145                 150                 155
caa aag gaa cgc aga gca act cga aga tgg atc aaa gcc atg tgt ttg        650
Gln Lys Glu Arg Arg Ala Thr Arg Arg Trp Ile Lys Ala Met Cys Leu
            160                 165                 170
gta aag aat tta gaa gca aat agt gta tta cct aaa ggt aca tca gaa        698
Val Lys Asn Leu Glu Ala Asn Ser Val Leu Pro Lys Gly Thr Ser Glu
        175                 180                 185
cac atg tta cca act gcc tta agc agt ctt cct ttg gaa gat ttt aag        746
His Met Leu Pro Thr Ala Leu Ser Ser Leu Pro Leu Glu Asp Phe Lys
        190                 195                 200
atc ctt gaa caa gat caa caa gat aaa aca ctg cta agt cta aat cta        794
Ile Leu Glu Gln Asp Gln Gln Asp Lys Thr Leu Leu Ser Leu Asn Leu
205                 210                 215                 220
aaa cag acc aag agt acc ttc att taaatttagc ttgcacagag cttgatgcct       848
Lys Gln Thr Lys Ser Thr Phe Ile
                225
atccttcatt cttttcagaa gtaaagataa ttatggcact tatgccaaaa ttcattattt       908
aataaagttt tacttgaagt aacattactg aatttgtgaa gacttgatta caaaagaata       968
aaaaacttca tatggaaatt ttatttgaaa atgagtggaa gcgccttaca ttagaattac      1028
ggacttaaaa attttgctaa taaattgtgt gtttgaaagg tgaaaaaaaa aaaaaa         1084
```

<210> 60
<211> 419
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 33..290

<221> sig_peptide
<222> 33..92

<223> Von Heijne matrix
      score 5.4
      seq WFVHSSALGLVLA/PP

<400> 60
```
aatggtaggc cttcatgtga gccagttact ac atg aat ctt cat ttc cca cag         53
                                     Met Asn Leu His Phe Pro Gln
                                     -20               -15

tgg ttt gtt cat tca tca gcg tta ggc ttg gtc ctg gct cca cct ttc        101
Trp Phe Val His Ser Ser Ala Leu Gly Leu Val Leu Ala Pro Pro Phe
            -10               -5                    1

tcc tct ccg ggc act gac ccc acc ttt ccg tgt att tac tgt agg cta        149
Ser Ser Pro Gly Thr Asp Pro Thr Phe Pro Cys Ile Tyr Cys Arg Leu
5                   10                  15

tta aat atg atc atg acc cgc ctt gca ttt tca ttc atc acc tgt tta        197
Leu Asn Met Ile Met Thr Arg Leu Ala Phe Ser Phe Ile Thr Cys Leu
20                  25                  30                  35

tgc cca aat tta aag gaa gtt tgt ctc att ttg cca gaa aaa aat tgt        245
Cys Pro Asn Leu Lys Glu Val Cys Leu Ile Leu Pro Glu Lys Asn Cys
                40                  45                  50

aat agt cgg cac gct gga ttt gta ggg cca gca aaa ttg cgg cag            290
Asn Ser Arg His Ala Gly Phe Val Gly Pro Ala Lys Leu Arg Gln
            55                  60                  65

tgaaactagt ttcacttcta aagcccttca tttcccacaa ggttaagctc tcgaaacccc      350
atttgatcct tggttcctat ttcgatcctc ctttggaatc tgaaaatcgg tctccatgtt      410
gtatgcaaa                                                              419
```

<210> 61
<211> 682
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 485..616

<221> polyA_site
<222> 669..682

<400> 61
```
ctcctttctc attccttatc ttgcgtgttt ttaccttttt ttcataacta agttttttgag      60
gaagttagtg ttcttttcaa agaaccggtt cgaaatgtac ttttctttgc tacttttttgt     120
tattttattg atcacatctt taatcttttg ttctctatac gtggcctgtt ttgatttatt     180
ttactattct tgctttctaa ggtaagtatt ttgttgtgta gtgctttatt ttttttcatct     240
ttcttcttga ataataatga catttttagg ttataaattt tcctctggta ctcagtttgc     300
ctcattaatt ttggcagtaa gcattctcct tttattgctt tctatgtagt ctttaatttt     360
gcttttaact tcttctttga tctaaggatt acctacttgt taattccaa atattatctt      420
atctatctat ctatctatct atctatctat ctatctatct acctatgtga gacgaagtct     480
ggct atg tcg ccg agg ctg gag tgc agt ggt gca atc ttg gct cac tgc       529
     Met Ser Pro Arg Leu Glu Cys Ser Gly Ala Ile Leu Ala His Cys
     1               5                   10                  15

aac ccc cgc ctc cca ggt tca agt tat tct cct gcc tca gct act tgg        577
Asn Pro Arg Leu Pro Gly Ser Ser Tyr Ser Pro Ala Ser Ala Thr Trp
            20                  25                  30

gtg aga gga tcc ctt gag ccg ggg agg ttg agg ctg cag tgagccataa        626
Val Arg Gly Ser Leu Glu Pro Gly Arg Leu Arg Leu Gln
            35                  40

ccactactct ccagcctgga taacaaaagt gagactctga ccaaaaaaaa aaaaaa         682
```

<210> 62
<211> 1191
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 54..995

<221> sig_peptide
<222> 54..227
<223> Von Heijne matrix
        score 4.1
        seq LVHHCPTWQWATG/EE

<221> polyA_signal
<222> 1130..1135

<221> polyA_site
<222> 1181..1191

<400> 62

```
cacggctgca cttttccatcc cgtcgcgggg ccggccgcta ctccggcccc agg atg          56
                                                                   Met

cag aat gtg att aat act gtg aag gga aag gca ctg gaa gtg gct gag          104
Gln Asn Val Ile Asn Thr Val Lys Gly Lys Ala Leu Glu Val Ala Glu
        -55               -50               -45

tac ctg acc ccg gtc ctc aag gaa tca aag ttt agg gaa aca ggt gta          152
Tyr Leu Thr Pro Val Leu Lys Glu Ser Lys Phe Arg Glu Thr Gly Val
        -40               -35               -30

att acc cca gaa gag ttt gtg gca gct gga gat cac cta gtc cac cac          200
Ile Thr Pro Glu Glu Phe Val Ala Ala Gly Asp His Leu Val His His
-25               -20               -15               -10

tgt cca aca tgg caa tgg gct aca ggg gaa gaa ttg aaa gtg aag gca          248
Cys Pro Thr Trp Gln Trp Ala Thr Gly Glu Glu Leu Lys Val Lys Ala
                  -5                1                 5

tac cta cca aca ggc aaa caa ttt ttg gta acc aaa aat gtg ccg tgc          296
Tyr Leu Pro Thr Gly Lys Gln Phe Leu Val Thr Lys Asn Val Pro Cys
        10                15                20

tat aag cgg tgc aaa cag atg gaa tat tca gat gaa ttg gaa gct atc          344
Tyr Lys Arg Cys Lys Gln Met Glu Tyr Ser Asp Glu Leu Glu Ala Ile
        25                30                35

att gaa gaa gat gat ggt gat ggc gga tgg gta gat aca tat cac aac          392
Ile Glu Glu Asp Asp Gly Asp Gly Gly Trp Val Asp Thr Tyr His Asn
40                45                50                55

aca ggt att aca gga ata acg gaa gcc gtt aaa gag atc aca ctg gaa          440
Thr Gly Ile Thr Gly Ile Thr Glu Ala Val Lys Glu Ile Thr Leu Glu
                  60                65                70

aat aag gac aat ata agg ctt caa gat tgc tca gca cta tgt gaa gag          488
Asn Lys Asp Asn Ile Arg Leu Gln Asp Cys Ser Ala Leu Cys Glu Glu
        75                80                85

gaa gaa gat gaa gat gaa gga gaa gct gca gat atg gaa gaa tat gaa          536
Glu Glu Asp Glu Asp Glu Gly Glu Ala Ala Asp Met Glu Glu Tyr Glu
        90                95                100

gag agt gga ttg ttg gaa aca gat gag gct acc cta gat aca agg aaa          584
Glu Ser Gly Leu Leu Glu Thr Asp Glu Ala Thr Leu Asp Thr Arg Lys
        105               110               115

ata gta gaa gct tgt aaa gcc aaa act gat gct ggc ggt gaa gat gct          632
Ile Val Glu Ala Cys Lys Ala Lys Thr Asp Ala Gly Gly Glu Asp Ala
120               125               130               135

att ttg caa acc aga act tat gac ctt tac atc act tat gat aaa tat          680
```

```
Ile Leu Gln Thr Arg Thr Tyr Asp Leu Tyr Ile Thr Tyr Asp Lys Tyr
            140                     145                     150
tac cag act cca cga tta tgg ttg ttt ggc tat gat gag caa cgg cag        728
Tyr Gln Thr Pro Arg Leu Trp Leu Phe Gly Tyr Asp Glu Gln Arg Gln
            155                     160                     165
cct tta aca gtt gag cac atg tat gaa gac atc agt cag gat cat gtg        776
Pro Leu Thr Val Glu His Met Tyr Glu Asp Ile Ser Gln Asp His Val
            170                     175                     180
aag aaa aca gtg acc att gaa aat cat cct cat ctg cca cca cct ccc        824
Lys Lys Thr Val Thr Ile Glu Asn His Pro His Leu Pro Pro Pro Pro
        185                     190                     195
atg tgt tca gtt cac cca tgc agg cat gct gag gtg atg aag aaa atc        872
Met Cys Ser Val His Pro Cys Arg His Ala Glu Val Met Lys Lys Ile
200                 205                     210                     215
att gag act gtt gca gaa gga ggg gga gaa ctt gga gtt cat atg tat        920
Ile Glu Thr Val Ala Glu Gly Gly Gly Glu Leu Gly Val His Met Tyr
                220                     225                     230
ctt ctt att ttc ttg aaa ttt gta caa gct gtc att cca aca ata gaa        968
Leu Leu Ile Phe Leu Lys Phe Val Gln Ala Val Ile Pro Thr Ile Glu
            235                     240                     245
tat gac tac aca aga cac ttc aca atg taatgaagag agcataaaat             1015
Tyr Asp Tyr Thr Arg His Phe Thr Met
            250                     255
ctatcctaat tattggttct gatttttaaa gaattaaccc atagatgtga ccattgacca     1075
tattcatcaa tatatacagt ttctctaata agggacttat atgtttatgc attaaataaa     1135
aatatgttcc actaccagcc ttacttgttt aataaaaatc agtgcaaaaa aaaaaa         1191
```

```
<210> 63
<211> 1008
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 657..923

<221> sig_peptide
<222> 657..896
<223> Von Heijne matrix
      score 3.5
      seq RGLLSACAPWGDG/ST

<221> polyA_signal
<222> 957..962

<221> polyA_site
<222> 974..1008

<400> 63
ntcgnatgtg gcacaaaacc cctctgctgg ctcatgtgtg caactgagac tgtcagagca      60
tggctagctc tggggtccag ctctgctggg tggggggctag agaggaagca gggagtatct    120
gcacacagga tgcctgcgct caggtggttg cagaagtcag tgcccaggcc cccccacaca     180
gtccccaaag gtccggcctc cccagcgcgg ggctcctcgt ttgaggggag gtgacttccc     240
tcccagcagg ctcttggaca cagtaagctt ccccagccct gcctgagcag cctttcctcc     300
ttgccctgtt ccccacctcc cggctccagt ccagggagct cccagggaag tggtcgaccc     360
ctccagtggc tgggccactc tgctagagtc catccgccaa gctgggggca tcggcaaggc     420
caagctgcgc agcatgaagg agcgaaagct ggagaagaag aagcagaagg agcaggagca    480
agtgagagcc acgagccaag gtgggcactt gatgtcggat ctcttcaaca agctggtcat    540
gaggcgcaag ggcatctctg ggaaagaacc tggggctggt gagggggccccg gaggagcctt    600
tgcccgcgctg tcagactcca tccctcctct gccgccaccg cagcagccac aggtag atg    659
```

```
                                                                       Met
                                                                       -80
agg aca agg acg act ggg aat cct agg ggg ctc cat gac acc ttc ccc        707
Arg Thr Arg Thr Thr Gly Asn Pro Arg Gly Leu His Asp Thr Phe Pro
              -75                    -70                    -65
cgc aga ccc aga ctt ggc cgt tgc tct gac atg gac aca gcc agg aca        755
Arg Arg Pro Arg Leu Gly Arg Cys Ser Asp Met Asp Thr Ala Arg Thr
              -60                    -55                    -50
agc tgc tca gac ctg ctt ccc tgg gag ggg gtg acg gaa cca gca ctg        803
Ser Cys Ser Asp Leu Leu Pro Trp Glu Gly Val Thr Glu Pro Ala Leu
              -45                    -40                    -35
tgt gga gac cag ctt caa gga acg gaa ggc tgg ctt gag gcc aca cag        851
Cys Gly Asp Gln Leu Gln Gly Thr Glu Gly Trp Leu Glu Ala Thr Gln
              -30                    -25                    -20
ctg ggg cgg gga ctt ctg tct gcc tgt gct cca tgg ggg gac ggc tcc        899
Leu Gly Arg Gly Leu Leu Ser Ala Cys Ala Pro Trp Gly Asp Gly Ser
              -15                    -10                    -5                    1
acc cag cct gtg cca ctg tgt tct taagaggctt ccagagaaaa cggcacacca        953
Thr Gln Pro Val Pro Leu Cys Ser
               5
atcaataaag aactgagcag aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaan          1008
```

```
<210> 64
<211> 568
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 18..311

<221> sig_peptide
<222> 18..62
<223> Von Heijne matrix
      score 8.4
      seq AMWLLCVALAVLA/WG

<400> 64
agtgctgctt acccatc atg gaa gca atg tgg ctc ctg tgt gtg gcg ttg         50
                    Met Glu Ala Met Trp Leu Leu Cys Val Ala Leu
                    -15                  -10             -5
gcg gtc ttg gca tgg ggc ttc ctc tgg gtt tgg gac tcc tca gaa cga        98
Ala Val Leu Ala Trp Gly Phe Leu Trp Val Trp Asp Ser Ser Glu Arg
              1               5                   10
atg aag agt cgg gag cag gga gga cgg ctg gga gcc gaa agc cgg acc       146
Met Lys Ser Arg Glu Gln Gly Gly Arg Leu Gly Ala Glu Ser Arg Thr
              15                  20                  25
ctg ctg gtc ata gcg cac cct gac gat gaa gcc atg ttt ttt gct ccc       194
Leu Leu Val Ile Ala His Pro Asp Asp Glu Ala Met Phe Phe Ala Pro
              30                  35                  40
aca gtg cta ggc ttg gcc cgc cta agg cac tgg gtg tac ctg ctt tgc       242
Thr Val Leu Gly Leu Ala Arg Leu Arg His Trp Val Tyr Leu Leu Cys
45                  50                  55                  60
ttc tct gca gtt ttc cgt agg gag cta agt gaa tac acc gaa ggt ctt       290
Phe Ser Ala Val Phe Arg Arg Glu Leu Ser Glu Tyr Thr Glu Gly Leu
              65                  70                  75
acc tct gaa ccc ctc aca gcc tagggacagg agcggccggc ttacctggtg          341
Thr Ser Glu Pro Leu Thr Ala
              80
ggttggggga cgtcggcagc tcgcgtacta cgccagcagg attgaggagc agagaaacag     401
```

```
ttgcagttgg ttgtattcag tacctgcatt tccgttggga actccacctg tacttgttat      461
tctgtggaac tttttttatt tgtagaagga gcaagaatat tgaccttact atatagcaca      521
cgaaacaatc tatgctgtat cgtgcctgct caatccttaa agttaac                    568
```

```
<210> 65
<211> 538
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 151..426

<221> sig_peptide
<222> 151..258
<223> Von Heijne matrix
      score 5.2
      seq KVALAGLLGFGLG/KV

<221> polyA_signal
<222> 505..510

<221> polyA_site
<222> 527..538

<400> 65
cactgggtca aggagtaagc agaggataaa caactggaag gagagcaagc acaaagtcat       60
catggcttca gcgtctgctc gtggaaacca agataaagat gcccattttc caccaccaag      120
caagcagctc tgcctttttc tcttgtaagc atg ctt gtc acc cag gga cta gtc      174
                                   Met Leu Val Thr Gln Gly Leu Val
                                   -35                     -30
tac caa ggt tat ttg gca gct aat tct aga ttt gga tca ttg ccc aaa      222
Tyr Gln Gly Tyr Leu Ala Ala Asn Ser Arg Phe Gly Ser Leu Pro Lys
        -25                 -20                 -15
gtt gca ctt gct ggt ctc ttg gga ttt ggc ctt gga aag gta tca tac      270
Val Ala Leu Ala Gly Leu Leu Gly Phe Gly Leu Gly Lys Val Ser Tyr
        -10                 -5                  1
ata gga gta tgc cag agt aaa ttc cat ttt ttt gaa gat cag ctc cgt      318
Ile Gly Val Cys Gln Ser Lys Phe His Phe Phe Glu Asp Gln Leu Arg
5                   10                  15                  20
ggg gct ggt ttt ggt cca cag cat aac agg cac tgc ctc ctt acc tgt      366
Gly Ala Gly Phe Gly Pro Gln His Asn Arg His Cys Leu Leu Thr Cys
                25                  30                  35
gag gaa tgc aaa ata aag cat gga tta agt gag aag gga gac tct cag      414
Glu Glu Cys Lys Ile Lys His Gly Leu Ser Glu Lys Gly Asp Ser Gln
            40                  45                  50
cct tca gct tcc taaattctgt gtctgtgact ttcgaagttt tttaaacctc          466
Pro Ser Ala Ser
            55
tgaatttgta cacatttaaa attcaagtg tactttaaaa taaaatactt ctaatggaac      526
aaaaaaaaaa aa                                                         538
```

```
<210> 66
<211> 1747
<212> DNA
<213> Homo sapiens

<220>
```

```
<221> CDS
<222> 10..1062

<221> sig_peptide
<222> 10..57
<223> Von Heijne matrix
      score 4.9
      seq FIYLQAHFTLCSG/WS

<221> polyA_signal
<222> 1710..1715

<221> polyA_site
<222> 1735..1747


<400> 66
gcctcacca atg gtt ccc ttc atc tat ctg caa gcc cac ttt aca ctc tgt          51
          Met Val Pro Phe Ile Tyr Leu Gln Ala His Phe Thr Leu Cys
              -15                 -10                 -5

tct ggg tgg tcc agc aca tac cgg gac ctc cgg aag ggt gtg tat gtg          99
Ser Gly Trp Ser Ser Thr Tyr Arg Asp Leu Arg Lys Gly Val Tyr Val
        1               5                   10

ccc tac acc cag ggc aag tgg gaa ggg gag ctg ggc acc gac ctg gta         147
Pro Tyr Thr Gln Gly Lys Trp Glu Gly Glu Leu Gly Thr Asp Leu Val
15                  20                  25                  30

agc atc ccc cat ggc ccc aac gtc act gtg cgt gcc aac att gct gcc         195
Ser Ile Pro His Gly Pro Asn Val Thr Val Arg Ala Asn Ile Ala Ala
                35                  40                  45

atc act gaa tca gac aag ttc ttc atc aac ggc tcc aac tgg gaa ggc         243
Ile Thr Glu Ser Asp Lys Phe Phe Ile Asn Gly Ser Asn Trp Glu Gly
                50                  55                  60

atc ctg ggg ctg gcc tat gct gag att gcc agg cct gac gac tcc ccg         291
Ile Leu Gly Leu Ala Tyr Ala Glu Ile Ala Arg Pro Asp Asp Ser Pro
            65                  70                  75

gag cct ttc ttt gac tct ctg gta aag cag acc cac gtt ccc aac ctc         339
Glu Pro Phe Phe Asp Ser Leu Val Lys Gln Thr His Val Pro Asn Leu
        80                  85                  90

ttc tcc ctg cag ctt tgt ggt gct ggc ttc ccc ctc aac cag tct gaa         387
Phe Ser Leu Gln Leu Cys Gly Ala Gly Phe Pro Leu Asn Gln Ser Glu
95                  100                 105                 110

gtg ctg gcc tct gtc gga ggg agc atg atc att gga ggt atc gac cac         435
Val Leu Ala Ser Val Gly Gly Ser Met Ile Ile Gly Gly Ile Asp His
                115                 120                 125

tcg ctg tac aca ggc agt ctc tgg tat aca ccc atc cgg cgg gag tgg         483
Ser Leu Tyr Thr Gly Ser Leu Trp Tyr Thr Pro Ile Arg Arg Glu Trp
            130                 135                 140

tat tat gag gtg atc att gtg cgg gtg gag atc aat gga cag gat ctg         531
Tyr Tyr Glu Val Ile Ile Val Arg Val Glu Ile Asn Gly Gln Asp Leu
            145                 150                 155

aaa atg gac tgc aag gag tac aac tat gac aag agc att gtg gac agt         579
Lys Met Asp Cys Lys Glu Tyr Asn Tyr Asp Lys Ser Ile Val Asp Ser
        160                 165                 170

ggc acc acc aac ctt cgt ttg ccc aag aaa gtg ttt gaa gct gca gtc         627
Gly Thr Thr Asn Leu Arg Leu Pro Lys Lys Val Phe Glu Ala Ala Val
175                 180                 185                 190

aaa tcc atc aag gca gcc tcc tcc acg gag aag ttc cct gac ggt ttc         675
Lys Ser Ile Lys Ala Ala Ser Ser Thr Glu Lys Phe Pro Asp Gly Phe
                195                 200                 205

tgg cta gga gag cag ctg gtg tgc tgg caa gca ggc acc acc cct tgg         723
Trp Leu Gly Glu Gln Leu Val Cys Trp Gln Ala Gly Thr Thr Pro Trp
            210                 215                 220

aac att ttc cca gtc atc tca ctc tac cta atg ggt gag gtt acc aac         771
```

```
Asn Ile Phe Pro Val Ile Ser Leu Tyr Leu Met Gly Glu Val Thr Asn
        225                 230                 235
cag tcc ttc cgc atc acc atc ctt ccg cag caa tac ctg cgg cca gtg         819
Gln Ser Phe Arg Ile Thr Ile Leu Pro Gln Gln Tyr Leu Arg Pro Val
        240                 245                 250
gaa gat gtg gcc acg tcc caa gac gac tgt tac aag ttt gcc atc tca         867
Glu Asp Val Ala Thr Ser Gln Asp Asp Cys Tyr Lys Phe Ala Ile Ser
255                 260                 265                 270
cag tca tcc acg ggc act gtt atg gga gct gtt atc atg gag ggc ttc         915
Gln Ser Ser Thr Gly Thr Val Met Gly Ala Val Ile Met Glu Gly Phe
                275                 280                 285
tac gtt gtc ttt gat cgg gcc cga aaa cga att ggc ttt gct gtc agc         963
Tyr Val Val Phe Asp Arg Ala Arg Lys Arg Ile Gly Phe Ala Val Ser
                290                 295                 300
gct tgc cat gtg cac gat gag ttc agg acg gca gcg gtg gaa ggc ccn        1011
Ala Cys His Val His Asp Glu Phe Arg Thr Ala Ala Val Glu Gly Pro
        305                 310                 315
ttt tgt cac ctt gga cat gga aga ctg tgg cta caa cat tcc aca gac        1059
Phe Cys His Leu Gly His Gly Arg Leu Trp Leu Gln His Ser Thr Asp
        320                 325                 330
aga tgagtcaacc ctcatgacca tagcctatgt catggctgcc atctgcgccc            1112
Arg
335
tcttcatgct gccactctgc ctcatggtgt gtcagtggcg ctgcctccgc tgcctgcgcc      1172
agcagcatga tgactttgct gatgacatct ccctgctgaa gtgaggaggc ccatgggcag      1232
aagataggga ttcccctgga ccacacctcc gtggttcact ttggtcacaa gtaggagaca      1292
cagatggcac ctgtggccag agcacctcag gaccctcccc acccaccaaa tgcctctgcc      1352
ttgatggaga aggaaaaggc tggcaaggtg ggttccaggg actgtacctg taggagacag      1412
aaaagagaag aaagaagcac tctgctggcg ggaatactct tggtcacctc aaatttaagt      1472
cgggaaattc tgctgcttga aacttcagcc ctgaaccttt gtcaccattc ctttaaattc      1532
tccaacccaa agtattcttc ttttcttagt ttcagaagta ctggcatcac acgcaggtta      1592
ccttggcgtg tgtccctgtg gtaccctggc agagaagaga ccaagcttgt ttccctgctg      1652
gccaaagtca gtaggagagg atgcacagtt tgctatttgc tttagagaca gggactgtat      1712
aaacaagcct aacattggtg caaaaaaaaa aaaa                                  1747
```

<210> 67
<211> 1686
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 78..491

<221> sig_peptide
<222> 78..218
<223> Von Heijne matrix
      score 5.8
      seq LMCFGALIGLCAC/IC

<221> polyA_signal
<222> 1652..1657

<221> polyA_site
<222> 1673..1686

<400> 67
```
ggtatagccc accagaaagg acagagtcat ttgatgtggt cacaaaatgt gtgagtttca       60
cactaactga gcagttc atg gag aaa ttt gtt gat ccc gga aac cac aat         110
                Met Glu Lys Phe Val Asp Pro Gly Asn His Asn
```

```
                              -45                    -40
agc ggg att gat ctc ctt agg acc tat ctt tgg cgt tgc cag ttc ctt        158
Ser Gly Ile Asp Leu Leu Arg Thr Tyr Leu Trp Arg Cys Gln Phe Leu
    -35                    -30                    -25
tta cct ttt gtg agt tta ggt ttg atg tgc ttt ggg gct ttg atc gga        206
Leu Pro Phe Val Ser Leu Gly Leu Met Cys Phe Gly Ala Leu Ile Gly
-20                    -15                    -10                    -5
ctt tgt gct tgc att tgc cga agc tta tat ccc acc att gcc acg ggc        254
Leu Cys Ala Cys Ile Cys Arg Ser Leu Tyr Pro Thr Ile Ala Thr Gly
                1                    5                    10
att ctc cat ctc ctt gca ggt ctg tgt aca ctg ggc tca gta agt tgt        302
Ile Leu His Leu Leu Ala Gly Leu Cys Thr Leu Gly Ser Val Ser Cys
            15                    20                    25
tat gtt gct gga att gaa cta ctc cac cag aaa cta gag ctc cct gac        350
Tyr Val Ala Gly Ile Glu Leu Leu His Gln Lys Leu Glu Leu Pro Asp
        30                    35                    40
aat gta tcc ggt gaa ttt gga tgg tcc ttc tgc ctt gct tgt gtc tct        398
Asn Val Ser Gly Glu Phe Gly Trp Ser Phe Cys Leu Ala Cys Val Ser
45                    50                    55                    60
gct ccc tta cag ttc atg gct tct gct ctc ttc atc tgg gct gct cac        446
Ala Pro Leu Gln Phe Met Ala Ser Ala Leu Phe Ile Trp Ala Ala His
                65                    70                    75
acc aac cgg aga gag tac acc tta atg aag gca tat cgt gtg gca            491
Thr Asn Arg Arg Glu Tyr Thr Leu Met Lys Ala Tyr Arg Val Ala
            80                    85                    90
tgagcaagaa actgcctgct ttacaattgc catttttatt tttttaaaat aatactgata       551
ttttccccac ctctcaattg tttttaattt ttatttgtgg atataccatt ttattatgaa       611
aatctatttt atttatacac attcaccact aaatacacac ttaataccac taaaatttat       671
gtggtttact ttaagcgatg ccatctttca aataaactaa tctaggtcta gacagaaaga       731
aatggataga gacttgacac aaatttatga aagaaaattg ggagtaggaa tgtgaccgaa       791
aacaagttgt gctaatgtct gttagacttt tcagtaaaac caaagtaact gtatctgttc       851
aactaaaaac tctatattag tttctttggg aaacctctca tcgtcaaaac tttatgttca       911
ctttgctgtt gtagatagcc agtcaaccag cagtattagt gctgttttca aagatttaag       971
ctctataaaa ttgggaaatt atctaagatc attttcccta agcattgaca catagcttca      1031
tctgaggtga gatatggcag ctgtttgtat ctgcactgtg tctgtctaca aagagtgaaa      1091
aatacagtgt ttacttgaaa ttttaacttt gtaactgcaa gaattccagt tcggccgggc      1151
gaggattagt attattttta actctccgta agattttcag taccaccaaa ttgttttgga      1211
ttttttttct ttcctcttca cataccaggg ttattaaaag tgtgctttct ttttacatta      1271
tattacagtt acaaggtaaa attcctcaac tgctatttat ttattccagc ccagtactat      1331
aaagaacgtt tcaccataat gaccctccag agctgggaaa cctaccacaa gatctaaagt      1391
tctggctgtc cattaacctc caactatggt ctttatttct tgtggtaata tgatgtgcct      1451
ttccttgcct aaatcccttc ctggtgtgta tcaacattat ttaatgtctt ctaattcagt      1511
cattttttat aagtatgtct ataaacattg aactttaaaa aacttattta tttattccac      1571
tactgtagca attgacagat taaaaaaatg taacttcata atttcttacc ataacctcaa      1631
tgtctttttt aaaaaataaa attaaaaatg aaagagacc caaaaaaaaa aaaaa           1686
```

```
<210> 68
<211> 542
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 69..371

<221> sig_peptide
<222> 69..287
<223> Von Heijne matrix
      score 4
      seq AVGFLFWVIVLTS/WI
```

<221> polyA_signal
<222> 510..515

<221> polyA_site
<222> 530..542

<400> 68
```
tgttacttag ggtcaaggct tgggtcttgc cccgcaaacc cttgggacga cccggcccca      60
gcgcagct atg aac ctg gag cga gtg tcc aat gag gag aaa ttg aac ctg       110
         Met Asn Leu Glu Arg Val Ser Asn Glu Glu Lys Leu Asn Leu
                 -70                 -65                 -60
tgc cgg aag tac tac ctg ggg ggg ttt gct ttc ttg cct ttt ctc tgg       158
Cys Arg Lys Tyr Tyr Leu Gly Gly Phe Ala Phe Leu Pro Phe Leu Trp
        -55                 -50                 -45
ttg gtc aac atc ttc tgg ttc tac cga gag gcc ttc ctt gtc cca gcc       206
Leu Val Asn Ile Phe Trp Phe Tyr Arg Glu Ala Phe Leu Val Pro Ala
        -40                 -35                 -30
tac aca gaa cag agc caa atc aaa ggc tat gtc tgg cgc tca gct gtg       254
Tyr Thr Glu Gln Ser Gln Ile Lys Gly Tyr Val Trp Arg Ser Ala Val
        -25                 -20                 -15
ggc ttc ctc ttc tgg gtg ata gtg ctc acc tcc tgg atc acc atc ttc       302
Gly Phe Leu Phe Trp Val Ile Val Leu Thr Ser Trp Ile Thr Ile Phe
    -10                 -5                  1                   5
cag atc tac cgg ccc cgc tgg ggt gcc ctt ggg gac tac ctc tcc ttc       350
Gln Ile Tyr Arg Pro Arg Trp Gly Ala Leu Gly Asp Tyr Leu Ser Phe
                10                  15                  20
acc ata ccc ctg ggc acc ccc tgacaacttc tgcacatact ggggccctgc          401
Thr Ile Pro Leu Gly Thr Pro
                25
ttattctccc aggacaggct ccttaaagca gaggagcctg tcctgggagc cccttctcaa      461
actcctaaga cttgttctca tgtcccacgt tctctgctga catcccccaa taaaggaccc      521
taactttcaa aaaaaaaaaa a                                               542
```

<210> 69
<211> 1174
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 2..757

<221> sig_peptide
<222> 2..205
<223> Von Heijne matrix
      score 7.3
      seq LRLILSPLPGAQP/QQ

<221> polyA_site
<222> 1160..1174

<400> 69
```
g atg cct gag ggc ccc gag ctg cac ctg gcc agc cag ttt gtg aat gag      49
  Met Pro Glu Gly Pro Glu Leu His Leu Ala Ser Gln Phe Val Asn Glu
          -65                 -60                 -55
gcc tgc agg gcg ctg gtg ttc ggc ggc tgc gtg gag aag tcc tct gtc        97
Ala Cys Arg Ala Leu Val Phe Gly Gly Cys Val Glu Lys Ser Ser Val
        -50                 -45                 -40
agc cgc aac cct gag gtg ccc ttt gag agc agt gcc tac cgc atc tca       145
```

```
Ser Arg Asn Pro Glu Val Pro Phe Glu Ser Ser Ala Tyr Arg Ile Ser
    -35             -30                    -25
gct tca gcc cgc ggc aag gag ctg cgc ctg ata ctg agc cct ctg cct      193
Ala Ser Ala Arg Gly Lys Glu Leu Arg Leu Ile Leu Ser Pro Leu Pro
-20             -15                    -10                     -5
ggg gcc cag cct caa cag gag cca ctg gcc ctg gtc ttc cgc ttc ggc      241
Gly Ala Gln Pro Gln Gln Glu Pro Leu Ala Leu Val Phe Arg Phe Gly
                1               5               10
atg tcc ggc tct ttt cag ctg gtg ccc cgc gag gag ctg cca cgc cat      289
Met Ser Gly Ser Phe Gln Leu Val Pro Arg Glu Glu Leu Pro Arg His
        15              20              25
gcc cac ctg cgc ttt tac acg gcc ccg cct ggc ccc cgg ctc gcc cta      337
Ala His Leu Arg Phe Tyr Thr Ala Pro Pro Gly Pro Arg Leu Ala Leu
        30              35              40
tgt ttc gtg gac atc cgc cgg ttc ggc cgc tgg gac ctt ggg gga aag      385
Cys Phe Val Asp Ile Arg Arg Phe Gly Arg Trp Asp Leu Gly Gly Lys
45              50              55              60
tgg cag ccg ggc cgc ggg ccc tgt gtc ttg cag gag tac cag cag ttc      433
Trp Gln Pro Gly Arg Gly Pro Cys Val Leu Gln Glu Tyr Gln Gln Phe
                65              70              75
agg gag aat gtg cta cga aac cta gcg gat aag gcc ttt gac cgg ccc      481
Arg Glu Asn Val Leu Arg Asn Leu Ala Asp Lys Ala Phe Asp Arg Pro
                80              85              90
atc tgc gag gcc ctc ctg gac cag agg ttc ttc aat ggc att ggc aac      529
Ile Cys Glu Ala Leu Leu Asp Gln Arg Phe Phe Asn Gly Ile Gly Asn
        95              100             105
tat ctg cgg gca gag atc ctg tac cgg ctg aag atc ccc ccc ttt gag      577
Tyr Leu Arg Ala Glu Ile Leu Tyr Arg Leu Lys Ile Pro Pro Phe Glu
        110             115             120
aag gcc cgc tcg gtc ctg gag gcc ctg cag cag cac agg ccg agc ccg      625
Lys Ala Arg Ser Val Leu Glu Ala Leu Gln Gln His Arg Pro Ser Pro
125             130             135             140
gag ctg acc ctg agc cag aag ata agg acc aag ctg cag aat tca gac      673
Glu Leu Thr Leu Ser Gln Lys Ile Arg Thr Lys Leu Gln Asn Ser Asp
                145             150             155
ctg ctg gag cta tgt cac tca gtg ccc aag gaa gtg gtc cag ttg ggt      721
Leu Leu Glu Leu Cys His Ser Val Pro Lys Glu Val Val Gln Leu Gly
        160             165             170
gag gcc aaa gat ggc agc aac ctc tgc ttc agc aaa tgattgtgta          767
Glu Ala Lys Asp Gly Ser Asn Leu Cys Phe Ser Lys
        175             180
acctggggc acttgtcccc ctctggacct gattcaccga tttggaagtt tgtagccta      827
gctgatactc aatggactag gcctcctcac ttgtcaatag tgtttccagg ctgggcgcag      887
tggctcatgc ctgtggtccc ggcacttcgg gaggccgagt ggggtggctc acctgaggtc      947
aggagttcga gaccatcctg gccaacatgg tgaaacccca tctccactaa aatgcaaaaa     1007
attagccagg tgtggtggcg ggcacctgta gtctcagcta ctcgggagga tgaggcagga     1067
aaatcgcttg aacccaggag gtggaggttg cagttgagct gagatcgtgc cattgcactc     1127
cagcctgggc aacgagagca aaactccatc tcaaaaaaaa aaaaaaa                  1174
```

<223> Von Heijne matrix
      score 7.3
      seq LRLILSPLPGAQP/QQ

<221> polyA_signal
<222> 1248..1253

<221> polyA_site
<222> 1272..1285

<400> 70
```
g atg cct gag ggc ccc gag ctg cac ctg gcc agc cag ttt gtg aat gag        49
  Met Pro Glu Gly Pro Glu Leu His Leu Ala Ser Gln Phe Val Asn Glu
          -65                 -60                 -55
gcc tgc agg gcg ctg gtg ttc ggc ggc tgc gtg gag aag tcc tct gtc        97
Ala Cys Arg Ala Leu Val Phe Gly Gly Cys Val Glu Lys Ser Ser Val
        -50                 -45                 -40
agc cgc aac cct gag gtg ccc ttt gag agc agt gcc tac cgc atc tca        145
Ser Arg Asn Pro Glu Val Pro Phe Glu Ser Ser Ala Tyr Arg Ile Ser
        -35                 -30                 -25
gct tca gcc cgc ggc aag gag ctg cgc ctg ata ctg agc cct ctg cct        193
Ala Ser Ala Arg Gly Lys Glu Leu Arg Leu Ile Leu Ser Pro Leu Pro
-20                 -15                 -10                 -5
ggg gcc cag ccc caa cag gag cca ctg gcc ctg gtc ttc cgc ttc ggc        241
Gly Ala Gln Pro Gln Gln Glu Pro Leu Ala Leu Val Phe Arg Phe Gly
                1                   5                   10
atg tcc ggc tct ttt cag ctg gtg ccc cgc gag gag ctg cca cgc cat        289
Met Ser Gly Ser Phe Gln Leu Val Pro Arg Glu Glu Leu Pro Arg His
        15                  20                  25
gcc cac ctg cgc ttt tac acg gcc ccg cct ggc ccc cgg ctc gcc cta        337
Ala His Leu Arg Phe Tyr Thr Ala Pro Pro Gly Pro Arg Leu Ala Leu
        30                  35                  40
tgt ttc gtg gac atc cgc cgg ttc ggc cgc tgg gac ctt ggg gga aag        385
Cys Phe Val Asp Ile Arg Arg Phe Gly Arg Trp Asp Leu Gly Gly Lys
45                  50                  55                  60
tgg cag ccg ggc cgc ggg ccc tgt gtc ttg cag gag tac cag cag ttc        433
Trp Gln Pro Gly Arg Gly Pro Cys Val Leu Gln Glu Tyr Gln Gln Phe
                65                  70                  75
agg ctg aag atc ccc ccc ttt gag aag gcc cgc tcg gtc ctg gag gcc        481
Arg Leu Lys Ile Pro Pro Phe Glu Lys Ala Arg Ser Val Leu Glu Ala
                80                  85                  90
ctg cag cag cac agg ccg agc ccg gag ctg acc ctg agc cag aag ata        529
Leu Gln Gln His Arg Pro Ser Pro Glu Leu Thr Leu Ser Gln Lys Ile
        95                  100                 105
agg acc aag ctg cag aat cca gac ctg ctg gag cta tgt cac tca gtg        577
Arg Thr Lys Leu Gln Asn Pro Asp Leu Leu Glu Leu Cys His Ser Val
110                 115                 120
ccc aag gaa gtg gac cag ttg ggg ggc agg ggc tac ggg tca gag agc        625
Pro Lys Glu Val Asp Gln Leu Gly Gly Arg Gly Tyr Gly Ser Glu Ser
125                 130                 135                 140
ggg gag gag gac ttt gct gcc ttt cga gcc tgg ctg cgc tgc tat ggc        673
Gly Glu Glu Asp Phe Ala Ala Phe Arg Ala Trp Leu Arg Cys Tyr Gly
                145                 150                 155
atg cca ggc atg agc tcc ctg cag gac cgg cat ggc cgt acc atc tgg        721
Met Pro Gly Met Ser Ser Leu Gln Asp Arg His Gly Arg Thr Ile Trp
                160                 165                 170
ttc cag ggg gat cct gga ccg ttg gca ccc aaa ggg cgc aag tcc cgc        769
Phe Gln Gly Asp Pro Gly Pro Leu Ala Pro Lys Gly Arg Lys Ser Arg
                175                 180                 185
aaa aag aaa tcc aag gcc aca cag ctg agt cct gag gac aga gtg gag        817
Lys Lys Lys Ser Lys Ala Thr Gln Leu Ser Pro Glu Asp Arg Val Glu
        190                 195                 200
```

```
gac gct ttg cct cca agc aag gcc cct tcc aag aca cga agg gca aag      865
Asp Ala Leu Pro Pro Ser Lys Ala Pro Ser Lys Thr Arg Arg Ala Lys
205                     210                 215                 220
aga gac ctt cct aag agg act gca acc cag cgg cct gag ggg acc agc      913
Arg Asp Leu Pro Lys Arg Thr Ala Thr Gln Arg Pro Glu Gly Thr Ser
                    225                     230                 235
ctc cag cag gac cca gaa gct ccc aca gtg ccc aag aag ggg agg agg      961
Leu Gln Gln Asp Pro Glu Ala Pro Thr Val Pro Lys Lys Gly Arg Arg
                240                     245                 250
aag ggg cga cag gca gcc tct ggc cac tgc aga ccc cgg aag gtc aag     1009
Lys Gly Arg Gln Ala Ala Ser Gly His Cys Arg Pro Arg Lys Val Lys
            255                     260                 265
gct gac atc cca tcc ttg gaa cca gag ggg acc tca gcc tct            1051
Ala Asp Ile Pro Ser Leu Glu Pro Glu Gly Thr Ser Ala Ser
            270                     275                 280
tagcaggagg ctctccttgc ttgcactcac cctttcttat tgtcttgccc tgcatctggg   1111
ggtctgaatt tttgggagca ggcaatatct gaaggtgcaa acaggcccta cggctgttcc   1171
ctgcacaact ctcatggttt taattgtacc ccatcttcca catctttaaa gctcatgtga   1231
aaaatgctgc atttttaata aactgataca tttgaactcc aaaaaaaaaa aaaa        1285
```

<210> 71
<211> 1398
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 2..1171

<221> sig_peptide
<222> 2..205
<223> Von Heijne matrix
       score 7.3
       seq LRLILSPLPGAQP/QQ

<221> polyA_signal
<222> 1368..1373

<221> polyA_site
<222> 1386..1398

<400> 71
```
g atg cct gag ggc ccc gag ctg cac ctg gcc agc cag ttt gtg aat gag   49
  Met Pro Glu Gly Pro Glu Leu His Leu Ala Ser Gln Phe Val Asn Glu
            -65                 -60                 -55
gcc tgc agg gcg ctg gtg ttc ggc ggc tgc gtg gag aag tcc tct gtc     97
Ala Cys Arg Ala Leu Val Phe Gly Gly Cys Val Glu Lys Ser Ser Val
        -50                 -45                 -40
agc cgc aac cct gag gtg ccc ttt gag agc agt gcc tac cgc atc tca    145
Ser Arg Asn Pro Glu Val Pro Phe Glu Ser Ser Ala Tyr Arg Ile Ser
    -35                 -30                 -25
gct tca gcc cgc ggc aag gag ctg cgc ctg ata ctg agc cct ctg cct    193
Ala Ser Ala Arg Gly Lys Glu Leu Arg Leu Ile Leu Ser Pro Leu Pro
-20                 -15                 -10                 -5
ggg gcc cag ccc caa cag gag cca ctg gcc ctg gtc ttc cgc ttc ggc    241
Gly Ala Gln Pro Gln Gln Glu Pro Leu Ala Leu Val Phe Arg Phe Gly
                1                   5                   10
atg tcc ggc tct ttt cag ctg gtg ccc cgc gag gag ctg cca cgc cat    289
Met Ser Gly Ser Phe Gln Leu Val Pro Arg Glu Glu Leu Pro Arg His
            15                  20                  25
```

```
gcc cac ctg cgc ttt tac acg gcc ccg cct ggc ccc cgg ctc gcc cta      337
Ala His Leu Arg Phe Tyr Thr Ala Pro Pro Gly Pro Arg Leu Ala Leu
    30                      35                      40

tgt ttc gtg gac atc cgc cgg ttc ggc cgc tgg gac ctt ggg gga aag      385
Cys Phe Val Asp Ile Arg Arg Phe Gly Arg Trp Asp Leu Gly Gly Lys
45                      50                      55                  60

tgg cag ccg ggc cgc ggg ccc tgt gtc ttg cag gag tac cag cag ttc      433
Trp Gln Pro Gly Arg Gly Pro Cys Val Leu Gln Glu Tyr Gln Gln Phe
                    65                      70                  75

agg gag aat gtg cta cga aac cta gcg gat aag gcc ttt gac cgg ccc      481
Arg Glu Asn Val Leu Arg Asn Leu Ala Asp Lys Ala Phe Asp Arg Pro
                80                      85                  90

atc tgc gag gcc ctc ctg gac cag agg ttc ttc aat ggc att ggc aac      529
Ile Cys Glu Ala Leu Leu Asp Gln Arg Phe Phe Asn Gly Ile Gly Asn
            95                      100                 105

tat ctg cgg gca gag atc ctg tac cgg ctg aag atc ccc ccc ttt gag      577
Tyr Leu Arg Ala Glu Ile Leu Tyr Arg Leu Lys Ile Pro Pro Phe Glu
        110                     115                 120

aag gcc cgc tcg gtc ctg gag gcc ctg cag cag cac agg ccg agc ccg      625
Lys Ala Arg Ser Val Leu Glu Ala Leu Gln Gln His Arg Pro Ser Pro
125                     130                 135                 140

gag ctg acc ctg agc cag aag ata agg acc aag ctg cag aat cca gac      673
Glu Leu Thr Leu Ser Gln Lys Ile Arg Thr Lys Leu Gln Asn Pro Asp
                    145                 150                 155

ctg ctg gag cta tgt cac tca gtg ccc aag gaa gtg gtc cag ttg ggg      721
Leu Leu Glu Leu Cys His Ser Val Pro Lys Glu Val Val Gln Leu Gly
                160                     165                 170

ggc aga ggc tac ggg tca gag agc ggg gag gag gac ttt gct gcc ttt      769
Gly Arg Gly Tyr Gly Ser Glu Ser Gly Glu Glu Asp Phe Ala Ala Phe
            175                     180                 185

cga gcc tgg ctg cgc tgc tat ggc atg cca ggc atg agc tcc ctg cag      817
Arg Ala Trp Leu Arg Cys Tyr Gly Met Pro Gly Met Ser Ser Leu Gln
        190                     195                 200

gac cgg cat ggc cgt acc atc tgg ttc cag ggg gat cct gga ccg ttg      865
Asp Arg His Gly Arg Thr Ile Trp Phe Gln Gly Asp Pro Gly Pro Leu
205                     210                 215                 220

gca ccc aaa ggg cgc aag tcc cgc aaa aag aaa tcc aag gcc aca cag      913
Ala Pro Lys Gly Arg Lys Ser Arg Lys Lys Lys Ser Lys Ala Thr Gln
                    225                 230                 235

ctg agt cct gag gac aga gtg gag gac gct ttg cct ccg agc aag gcc      961
Leu Ser Pro Glu Asp Arg Val Glu Asp Ala Leu Pro Pro Ser Lys Ala
                240                     245                 250

cct tcc agg aca cga agg gca aag aga gac ctt cct aag agg act gca     1009
Pro Ser Arg Thr Arg Arg Ala Lys Arg Asp Leu Pro Lys Arg Thr Ala
            255                     260                 265

acc cag cgg cct gag ggg acc agc ctc cag cag gac cca gaa gct ccc     1057
Thr Gln Arg Pro Glu Gly Thr Ser Leu Gln Gln Asp Pro Glu Ala Pro
        270                     275                 280

aca gtg ccc aag aag ggg agg agg aag ggg cga cag gca gcc tct ggc     1105
Thr Val Pro Lys Lys Gly Arg Arg Lys Gly Arg Gln Ala Ala Ser Gly
285                     290                     295                 300

cac tgc aga ccc cgg aag gtc aag gct gac atc cca tcc ttg gaa cca     1153
His Cys Arg Pro Arg Lys Val Lys Ala Asp Ile Pro Ser Leu Glu Pro
                    305                 310                 315

gag ggg acc tca gcc tct tagcaggagg ctctccttgc ttgcactcac           1201
Glu Gly Thr Ser Ala Ser
                    320

cctttcttat tgtcttgccc tgcatctggg ggtctgaatt tttgggagca ggcaatatct  1261
gaaggtgcaa acaggcccta cggctgttcc ctgcacaact ctcatggttt taattgtacc  1321
ccatcttcca catctttaaa gctcatgtga aaaatgctgc atttttaata aactgataca  1381
tttgaaaaaa aaaaaaa                                                 1398
```

```
<210> 72
<211> 821
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 42..611

<221> sig_peptide
<222> 42..287
<223> Von Heijne matrix
      score 4.4
      seq NLPHLQVVGLTWG/HI

<221> polyA_signal
<222> 787..792

<221> polyA_site
<222> 808..821

<400> 72
ccgttgccag ttctgcgcgt gtcctgcgtc tccagtatgg a atg tat gtt tgg ccc    56
                                              Met Tyr Val Trp Pro
                                                            -80
tgt gct gtg gtc ctg gcc cag tac ctt tgg ttt cac aga aga tct ctg     104
Cys Ala Val Val Leu Ala Gln Tyr Leu Trp Phe His Arg Arg Ser Leu
        -75             -70                 -65
cca ggc aag gcc atc tta gag att gga gca gga gtg agc ctt cca gga     152
Pro Gly Lys Ala Ile Leu Glu Ile Gly Ala Gly Val Ser Leu Pro Gly
    -60             -55                 -50
att ttg act gcc aaa tgt ggt gca gaa gta ata ctg tca gac agc tca     200
Ile Leu Thr Ala Lys Cys Gly Ala Glu Val Ile Leu Ser Asp Ser Ser
-45             -40                 -35                 -30
gaa ctg cct cac tgt ctg gaa gtc tgt cgg caa agc tgc caa atg aat     248
Glu Leu Pro His Cys Leu Glu Val Cys Arg Gln Ser Cys Gln Met Asn
            -25                 -20                 -15
aac ctg cca cat ctg cag gtg gta gga cta aca tgg ggt cat ata tct     296
Asn Leu Pro His Leu Gln Val Val Gly Leu Thr Trp Gly His Ile Ser
        -10                 -5                  1
tgg gat ctt ctg gct cta cca cca caa gat att atc ctt gca tct gat     344
Trp Asp Leu Leu Ala Leu Pro Pro Gln Asp Ile Ile Leu Ala Ser Asp
5                   10                  15
gtg ttc ttt gaa cca gaa gat ttt gaa gac att ttg gct aca ata tat     392
Val Phe Phe Glu Pro Glu Asp Phe Glu Asp Ile Leu Ala Thr Ile Tyr
20                  25                  30                  35
ttt ttg atg cac aag aat ccc aag gtc caa ttg tgg tct act tat caa     440
Phe Leu Met His Lys Asn Pro Lys Val Gln Leu Trp Ser Thr Tyr Gln
            40                  45                  50
gtt agg agt gct gac tgg tca ctt gaa gct tta ctc tac aaa tgg gat     488
Val Arg Ser Ala Asp Trp Ser Leu Glu Ala Leu Leu Tyr Lys Trp Asp
            55                  60                  65
atg aaa tgt gtc cac att cct ctt gag tct ttt gat gca gac aaa gaa     536
Met Lys Cys Val His Ile Pro Leu Glu Ser Phe Asp Ala Asp Lys Glu
        70                  75                  80
gat ata gca gaa tct acc ctt cca gga aga cat aca gtt gaa atg ctg     584
Asp Ile Ala Glu Ser Thr Leu Pro Gly Arg His Thr Val Glu Met Leu
        85                  90                  95
gtc att tcc ctt gca aag gac agt ctc tgaattatac ctacaacctg          631
Val Ile Ser Phe Ala Lys Asp Ser Leu
```

```
           100                105
ttctgggaca gtatcaatac tgatgagcaa cctggcacac aaactatgag cagaccactt      691
cagcttgaga atgcagtggg tctgaagatg gtcaagtctg tctgccttag attttgatgt      751
cacctagaca acacttaaac tcatatgaaa caaaaattaa aatacgtatt acaagtaaaa      811
aaaaaaaaaa                                                            821
```

```
<210> 73
<211> 916
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 62..916

<221> sig_peptide
<222> 62..757
<223> Von Heijne matrix
      score 4.2
      seq LVTPAALRPLVLG/GN

<221> polyA_site
<222> 904..916

<400> 73
cctgaatgac ttgaatgttt ccccgcctga gctaacagtc catgtgggtg attcagctct      60
g atg gga tgt gtt ttc cag agc aca gaa gac aaa cgt ata ttc aag ata    109
  Met Gly Cys Val Phe Gln Ser Thr Glu Asp Lys Arg Ile Phe Lys Ile
          -230                -225                -220
gac tgg act ctg tca cca gga gag cac gcc aag gac gaa tat gtg cta      157
Asp Trp Thr Leu Ser Pro Gly Glu His Ala Lys Asp Glu Tyr Val Leu
        -215                -210                -205
tac tat tac tcc aat ctc agt gtg cct att ggg cgc ttc cag aac cgc      205
Tyr Tyr Tyr Ser Asn Leu Ser Val Pro Ile Gly Arg Phe Gln Asn Arg
-200                -195                -190                -185
gta cac ttg atg ggg gac aac tta tgc aat gat ggc tct ctc ctg ctc      253
Val His Leu Met Gly Asp Asn Leu Cys Asn Asp Gly Ser Leu Leu Leu
          -180                -175                -170
caa gat gtg caa gag gct gac cag gga acc tat atc tgt gaa atc cgc      301
Gln Asp Val Gln Glu Ala Asp Gln Gly Thr Tyr Ile Cys Glu Ile Arg
        -165                -160                -155
ctc aaa ggg gag agc cag gtg ttc aag aag gcg gtg gta ctg cat gtg      349
Leu Lys Gly Glu Ser Gln Val Phe Lys Lys Ala Val Val Leu His Val
        -150                -145                -140
ctt cca gag gag ccc aaa gag ctc atg gtc cat gtg ggt gga ttg att      397
Leu Pro Glu Glu Pro Lys Glu Leu Met Val His Val Gly Gly Leu Ile
        -135                -130                -125
cag atg gga tgt gtt ttc cag agc aca gaa gtg aaa cac gtg acc aag      445
Gln Met Gly Cys Val Phe Gln Ser Thr Glu Val Lys His Val Thr Lys
-120                -115                -110                -105
gta gaa tgg ata ttt tca gga cgg cgc gca aag gag gag att gta ttt      493
Val Glu Trp Ile Phe Ser Gly Arg Arg Ala Lys Glu Glu Ile Val Phe
                -100                -95                -90
cgt tac tac cac aaa ctc agg atg tct gcg gag tac tcc cag agc tgg      541
Arg Tyr Tyr His Lys Leu Arg Met Ser Ala Glu Tyr Ser Gln Ser Trp
        -85                -80                -75
ggc cac ttc cag aat cgt gtg aac ctg gtg ggg gac att ttc cgc aat      589
Gly His Phe Gln Asn Arg Val Asn Leu Val Gly Asp Ile Phe Arg Asn
        -70                 -65                 -60
gac ggt tcc atc atg ctt caa gga gtg agg gag tca gat gga gga aac      637
```

```
Asp Gly Ser Ile Met Leu Gln Gly Val Arg Glu Ser Asp Gly Gly Asn
-55              -50                    -45
tac acc tgc agt atc cac cta ggg aac ctg gtg ttc aag aaa acc att    685
Tyr Thr Cys Ser Ile His Leu Gly Asn Leu Val Phe Lys Lys Thr Ile
-40              -35              -30                  -25
gtg ctg cat gtc agc ccg gaa gag cct cga aca ctg gtg acc ccg gca    733
Val Leu His Val Ser Pro Glu Glu Pro Arg Thr Leu Val Thr Pro Ala
            -20              -15              -10
gcc ctg agg cct ctg gtc ttg ggt ggt aat cag ttg gtg atc att gtg    781
Ala Leu Arg Pro Leu Val Leu Gly Gly Asn Gln Leu Val Ile Ile Val
            -5              1              5
gga att gtc tgt gcc aca atc ctg ctg ctc cct gtc ctg ata ttg atc    829
Gly Ile Val Cys Ala Thr Ile Leu Leu Leu Pro Val Leu Ile Leu Ile
    10              15                  20
gtg aag aag acc tgt gga aat aag agt tca gtg aat tct aca gtc ttg    877
Val Lys Lys Thr Cys Gly Asn Lys Ser Ser Val Asn Ser Thr Val Leu
25                  30                  35                  40
gtg aag aac acg aag aag act aat cca aaa aaa aaa aaa                916
Val Lys Asn Thr Lys Lys Thr Asn Pro Lys Lys Lys Lys
                45                  50
```

```
<210> 74
<211> 1153
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 62..520

<221> polyA_signal
<222> 1124..1129

<221> polyA_site
<222> 1141..1153

<400> 74
cctgaatgac ttgaatgttt ccccgcctga gctaacagtc catgtgggtg attcagctct    60
g atg gga tgt gtt ttc cag agc aca gta gac aaa tgt ata ttc aag ata   109
  Met Gly Cys Val Phe Gln Ser Thr Val Asp Lys Cys Ile Phe Lys Ile
  1           5                   10                  15
gac tgg act ctg tca cca gga gag cac gcc aag gac gaa tat gtg cta    157
Asp Trp Thr Leu Ser Pro Gly Glu His Ala Lys Asp Glu Tyr Val Leu
        20              25                  30
tac tat tac tcc aat ctc agt gtg cct att ggg cgc ttc cag aac cgc    205
Tyr Tyr Tyr Ser Asn Leu Ser Val Pro Ile Gly Arg Phe Gln Asn Arg
        35              40                  45
gta cac ttg atg ggg gac atc tta tgc aat gat ggc tct ctc ctg ctc    253
Val His Leu Met Gly Asp Ile Leu Cys Asn Asp Gly Ser Leu Leu Leu
    50              55                  60
caa gat gtg caa gag gct gac cag gga acc tat atc tgt gaa atc cgc    301
Gln Asp Val Gln Glu Ala Asp Gln Gly Thr Tyr Ile Cys Glu Ile Arg
65              70                  75                  80
ctc aaa ggg gag agc cag gtg ttc aag aag gcg gtg gta ctg cat gtg    349
Leu Lys Gly Glu Ser Gln Val Phe Lys Lys Ala Val Val Leu His Val
            85                  90                  95
ctt cca gag gag ccc aaa gag ctc atg gtc cat gtg ggt gga ttg att    397
Leu Pro Glu Glu Pro Lys Glu Leu Met Val His Val Gly Gly Leu Ile
            100             105                 110
cag atg gga tgt gtt ttc cag agc aca gaa gtg aaa cac gtg acc aag    445
```

```
Gln Met Gly Cys Val Phe Gln Ser Thr Glu Val Lys His Val Thr Lys
        115                 120                 125
gta gaa tgg ata ttt tca gga cgg cgc gca aag gta aca agg agg aaa      493
Val Glu Trp Ile Phe Ser Gly Arg Arg Ala Lys Val Thr Arg Arg Lys
        130                 135                 140
cat cac tgt gtt aga gaa ggc tct ggc tgatggtatc aggacaaagg          540
His His Cys Val Arg Glu Gly Ser Gly
145                 150
tagaatcagg cacatgagga ggtgttgcaa gagcctgggc ttggtgctt atcagaactg    600
gaccttctcc tagcaatttc agctttctgg tgggaaaggt aactccaatg aagaacaaga    660
acaagaagat gatgatgatg cttaactttt tggatgccga tatgagattg tacatgtaaa    720
gcattttgta taagacttgg cccctgcatt ttagtttcct tctttctccc ttttccttcg    780
tatagagtcc atgggagaat gaggagagtg attttttgtgg cccagccaag aaagcaatgg   840
gctagacatt aaaatgatta cactttattt cttactgggg ttagttctgt gagttttcat    900
ctgtgcccca ttgccccatt tatgtgatgg agggaatttt catgggtact tcacgtgttg    960
ggattgattg atcctggggg ccagggtgaa gggtatttta cgggacctct ataaagcagg   1020
aagaagcaag tttattcttt agaccagtag ctctcaacca tgatgtggtc atatatttat   1080
gggtcaacat gtgttgtggg gatatcccaa gtaacttgtt attaataaaa gttaagttgc   1140
aaaaaaaaaa aaa                                                      1153
```

```
<210> 75
<211> 1517
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 21..167

<400> 75
ctctgaaatg cttgtctttt atg ctg gna ggt gac cat agg gct ctg ctt tta    53
                       Met Leu Xaa Gly Asp His Arg Ala Leu Leu Leu
                        1               5                   10
aag ata tgg ctg ctt caa agg cca gag tca cag gaa gga ctt ctt cca      101
Lys Ile Trp Leu Leu Gln Arg Pro Glu Ser Gln Glu Gly Leu Leu Pro
        15                  20                  25
ggg aga tta gtg gtg atg gag agg aga gtt aaa atg acc tca tgt cct      149
Gly Arg Leu Val Val Met Glu Arg Arg Val Lys Met Thr Ser Cys Pro
        30                  35                  40
tct tgt cca cgg ttt tgt tgagttttca ctcttctaat gcaagggtct            197
Ser Cys Pro Arg Phe Cys
        45
cacactgtga accacttagg atgtgatcac tttcaggtgg ccaggaatgt tgaatgtctt    257
tggctcagtt catttaaaaa agatatctat ttgaaagttc tcagagttgt acatatgttt    317
cacagtacag gatctgtaca taaaagtttc tttcctaaac cattcaccaa gagccaatat    377
ctaggcattt tcttggtagc acaaattttc ttattgctta gaaaattgtc ctccttgtta    437
tttctgtttg taagacttaa gtgagttagg tctttaagga aagcaacgct cctctgaaat    497
gcttgtcttt tatgctggga ggtgaccata gggctctgct ttaaagata tggctgcttc    557
aaaggccaga gtcacaggaa ggacttcttc cagggagatt agtggtgatg gagaggagag    617
ttaaaatgac ctcatgtcct tcttgtccac ggttttgttg agttttcact cttctaatgc    677
aagggtctca cacttaggat gtgatcactt tcaggtggcc aggaatgttg               737
aatgtctttg gctcagttca tttaaaaaag atatctattt gaaagttctc agagttgtac    797
atatgtttca gtacagga tctgtacata aaagtttctt ttctaaacca ttcaccaaga      857
gccaatatct aggcatttttc ttggtagcac aaattttctt attgcttaga aaattgtcct    917
ccttgttatt tctgtttgta agacttaagt gagttaggtc ttaaggaaa gcaacgctcc     977
tctgaaatgc ttgtctttna tgctgggagg tgaccatagg gctctgcttt aaagatatg    1037
gctgcttcaa aggccagagt cacaggaagg acttcttcca gggagattag tggtgatgga   1097
gaggagagtt aaaatgacct catgtccttc ttgtccacgg ttttgttgag ttttcactct   1157
tctaatgcaa gggtctcaca ctgtgaacca cttaggatgt gatcactttc aggtggccag   1217
gaatgttgaa tgtcttttggc tcagttcatt taaaaaagat ttctatttga aagttctcag   1277
```

```
agttgtacat atgtttcaca gtacaggatc tgtacataaa agttctttc ctaaaccatt        1337
caccaagagc caatatctag gcattttctt ggtagcacaa attttcttat tgcttagaaa        1397
attgtcctcc ttgttatttc tgtttgtaag acttaagtga gttaggtctt taaggaaagc        1457
aacgctcctc tgaaatgctt gtcttttatg ctgggaggtg accatagggc tctgctttta        1517
```

<210> 76
<211> 526
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 22..318

<221> sig_peptide
<222> 22..93
<223> Von Heijne matrix
      score 4.6
      seq FFIFCSLNTLLLG/GV

<221> polyA_signal
<222> 497..502

<221> polyA_site
<222> 516..526

<400> 76
```
ctgcctgctg cttgctgcac c atg aag tct gcc aag ctg gga ttt ctt cta         51
                        Met Lys Ser Ala Lys Leu Gly Phe Leu Leu
                                -20                         -15

aga ttc ttc atc ttc tgc tca ttg aat acc ctg tta ttg ggt ggt gtt         99
Arg Phe Phe Ile Phe Cys Ser Leu Asn Thr Leu Leu Leu Gly Gly Val
        -10                     -5                          1

aat aaa att gcg gag aag ata tgt gga gac ctc aaa gat ccc tgc aaa        147
Asn Lys Ile Ala Glu Lys Ile Cys Gly Asp Leu Lys Asp Pro Cys Lys
        5                   10                  15

ttg gac atg aat ttt gga agc tgc tat gaa gtt cac ttt aga tat ttc        195
Leu Asp Met Asn Phe Gly Ser Cys Tyr Glu Val His Phe Arg Tyr Phe
    20                  25                  30

tac aac aga acc tcc aaa aga tgt gaa act ttt gtc ttc tcc ggc tgt        243
Tyr Asn Arg Thr Ser Lys Arg Cys Glu Thr Phe Val Phe Ser Gly Cys
35                  40                  45                  50

aat ggc aac ctt aac aac ttc aag ctt aaa ata gaa cgt gaa gta gcc        291
Asn Gly Asn Leu Asn Asn Phe Lys Leu Lys Ile Glu Arg Glu Val Ala
                55                  60                  65

tgt gtt gca aaa tac aaa cca ccg agg tgagaggatg tgaactcatg            338
Cys Val Ala Lys Tyr Lys Pro Pro Arg
                70                  75

aagttgtctg ctgcaccatc cgaaataaag acacaagaaa attcagactg attttgaaat   398
ctttgtaata tttccataat gctttaagct tccatatgtt tgctattttc ctgaccctag   458
ttttgtcttt cctggaaatt aactgtatga tcattagaat gaaagagtct ttctgtcaaa   518
aaaaaaaa                                                            526
```

<210> 77
<211> 352
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 8..292

<221> sig_peptide
<222> 8..118
<223> Von Heijne matrix
       score 5.6
       seq WLLLDALLRLGDT/KK

<221> polyA_signal
<222> 317..322

<221> polyA_site
<222> 339..352

<400> 77
```
ctgagat atg gca agt ccc gct gta aac agg tgg aaa agg cca agg ttg              49
        Met Ala Ser Pro Ala Val Asn Arg Trp Lys Arg Pro Arg Leu
        -35                     -30                     -25
aag ccg gtg tgg cca cgg cgc ttg gaa tcc tgg ttg ttg ctg gat gct              97
Lys Pro Val Trp Pro Arg Arg Leu Glu Ser Trp Leu Leu Leu Asp Ala
        -20                     -15                     -10
ctt ttg cga tta gga gat acc aaa aaa aag cga cag cct gaa gca gcc             145
Leu Leu Arg Leu Gly Asp Thr Lys Lys Lys Arg Gln Pro Glu Ala Ala
        -5              1               5
aca aaa tcc tgt gtt aga agc agc tgt ggg ggt ccc agt gga gat ggg            193
Thr Lys Ser Cys Val Arg Ser Ser Cys Gly Gly Pro Ser Gly Asp Gly
10                  15                  20                  25
cct ccc cca tgc ctc cag cag cct gac cct cgt gcc ctg tct cag gcg            241
Pro Pro Pro Cys Leu Gln Gln Pro Asp Pro Arg Ala Leu Ser Gln Ala
                30                  35                  40
ttc tct aga tcc ttt cct ctg ttt ccc tct ctc gct ggc aaa agt atg            289
Phe Ser Arg Ser Phe Pro Leu Phe Pro Ser Leu Ala Gly Lys Ser Met
                45                  50                  55
atc taattgaaac aagactgaag gatcaataaa cagccatctg ccccttcaaa                  342
Ile
aaaaaaaaaa                                                                   352
```

<210> 78
<211> 542
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 16..378

<221> sig_peptide
<222> 16..84
<223> Von Heijne matrix
       score 9.8
       seq FLLFFFLFLLTRG/SL

<221> polyA_signal
<222> 502..507

<221> polyA_site
<222> 522..542

<400> 78
cacgacctgt gggcc atg atg cta ccc caa tgg ctg ctg ctg ctg ttc ctt   51
                 Met Met Leu Pro Gln Trp Leu Leu Leu Leu Phe Leu
                     -20   -15
ctc ttc ttc ttt ctc ttc ctc ctc acc agg ggc tca ctt tct cca aca   99
Leu Phe Phe Phe Leu Phe Leu Leu Thr Arg Gly Ser Leu Ser Pro Thr
    -10   -5   1   5
aaa tat aac ctt ttg gag ctc aag gag tct tgc atc cgg aac cag gac   147
Lys Tyr Asn Leu Leu Glu Leu Lys Glu Ser Cys Ile Arg Asn Gln Asp
                10   15   20
tgc gag act ggc tgc tgc caa cgt gct cca gac aat tgc gag tcg cac   195
Cys Glu Thr Gly Cys Cys Gln Arg Ala Pro Asp Asn Cys Glu Ser His
            25   30   35
tgc gcg gag aag ggg tcc gag ggc agt ctg tgt caa acg cag gtg ttc   243
Cys Ala Glu Lys Gly Ser Glu Gly Ser Leu Cys Gln Thr Gln Val Phe
        40   45   50
ttt ggc caa tat aga gcg tgt ccc tgc ctg cgg aac ctg act tgt ata   291
Phe Gly Gln Tyr Arg Ala Cys Pro Cys Leu Arg Asn Leu Thr Cys Ile
    55   60   65
tat tca aag aat gag aaa tgg ctt agc atc gcc tat ggc cgt tgt cag   339
Tyr Ser Lys Asn Glu Lys Trp Leu Ser Ile Ala Tyr Gly Arg Cys Gln
70   75   80   85
aaa att gga agg cag aag ttg gct aag aaa atg ttc ttc tagtgctccc   388
Lys Ile Gly Arg Gln Lys Leu Ala Lys Lys Met Phe Phe
                90   95
tccttcttgc tgcctcctcc tcctccacct gctctcctcc ctacccagag ctctgtgttc   448
accctgttcc ccagagcctc caccatgagt ggagggaagt ggggagtgat tgaaataaag   508
agcttttttca atgaaaaaaa aaaaaaaaaa aaaa   542

<210> 79
<211> 233
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 57..233

<400> 79
gcaaaaccaa aaccagcacc gatcccgaca tagatcagtg acgtcttttt cttcag atg   59
                                                             Met
                                                               1
atc cta tgt ttc ctt ctt cct cat cat cgt ctt cag gaa gcc aga cag   107
Ile Leu Cys Phe Leu Leu Pro His His Arg Leu Gln Glu Ala Arg Gln
          5   10   15
att caa gta ttg aag atg ctg cca agg gaa aaa tta aga aga aga gaa   155
Ile Gln Val Leu Lys Met Leu Pro Arg Glu Lys Leu Arg Arg Arg Glu
        20   25   30
gag aga aaa caa ata aat ggg aaa aaa gaa agg aca aaa tat gaa aca   203
Glu Arg Lys Gln Ile Asn Gly Lys Lys Glu Arg Thr Lys Tyr Glu Thr
        35   40   45
cca aga aaa aga gaa gga aaa aaa aaa aaa   233
Pro Arg Lys Arg Glu Gly Lys Lys Lys Lys
50   55

<210> 80
<211> 660
<212> DNA

<213> Homo sapiens

<220>
<221> CDS
<222> 83..340

<221> sig_peptide
<222> 83..124
<223> Von Heijne matrix
      score 7.5
      seq VALNLILVPCCAA/WC

<221> polyA_signal
<222> 573..578

<221> polyA_site
<222> 607..660

<400> 80
gaatttgtaa aacttctgct cgtttacact gcacattgaa tacaggtaac taattggaag   60
gagaggggag atcactcttt tg atg gtg gcc ctg aac ctc att ctg gtt ccc   112
                         Met Val Ala Leu Asn Leu Ile Leu Val Pro
                                                -10   -5

tgc tgc gct gct tgg tgt gac cca cgg agg atc cac tcc cag gat gac   160
Cys Cys Ala Ala Trp Cys Asp Pro Arg Arg Ile His Ser Gln Asp Asp
             1   5   10

gtg ccc cgt agc tct gct gct gat act ggg tct gcg atg cag cgg cgt   208
Val Pro Arg Ser Ser Ala Ala Asp Thr Gly Ser Ala Met Gln Arg Arg
         15   20   25

gag gcc tgg gct ggt tgg aga agg tca caa ccc ttc tct gtt ggt ctg   256
Glu Ala Trp Ala Gly Trp Arg Arg Ser Gln Pro Phe Ser Val Gly Leu
     30   35   40

cct tct gct gaa aga ctc gag aac caa cca ggg aag ctg tcc tgg agg   304
Pro Ser Ala Glu Arg Leu Glu Asn Gln Pro Gly Lys Leu Ser Trp Arg
45   50   55   60

tcc ctg gtc gga gag gga tat aga atc tgt gac ctc tgacaactgt   350
Ser Leu Val Gly Glu Gly Tyr Arg Ile Cys Asp Leu
                 65   70

gaagccaccc tgggctacag aaaccacagt cttcccagca attattacaa ttcttgaatt   410
ccttggggat tttttactgc cctttcaaag cacttaagtg ttagatctaa cgtgttccag   470
tgtctgtctg aggtgactta aaaaatcaga acaaaacttc tattatccag agtcatggga   530
gagtacaccc tttccaggaa taatgttttg ggaaacactg aaatgaaatc ttcccagtat   590
tataaattgt gtatttaaaa aaagaaactt ttctgaatgc ctacctggcg gtgtatacca   650
ggcagtgtgc   660

<210> 81
<211> 605
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 47..541

<221> sig_peptide
<222> 47..220
<223> Von Heijne matrix
      score 5.4
      seq QLLDSVLWLGALG/LT

<221> polyA_site
<222> 597..605

<400> 81
```
aaagtgggag gagcactagg tcttcccgtc acctccacct ctctcc atg acc cgg     55
                                                     Met Thr Arg
ctc tgc tta ccc aga ccc gaa gca cgt gag gat ccg atc cca gtt cct   103
Leu Cys Leu Pro Arg Pro Glu Ala Arg Glu Asp Pro Ile Pro Val Pro
-55                 -50                 -45                 -40
cca agg ggc ctg ggt gct ggg gag ggg tca ggt agt cca gtg cgt cca   151
Pro Arg Gly Leu Gly Ala Gly Glu Gly Ser Gly Ser Pro Val Arg Pro
                -35                 -30                 -25
cct gta tcc acc tgg ggc cct agc tgg gcc cag ctc ctg gac agt gtc   199
Pro Val Ser Thr Trp Gly Pro Ser Trp Ala Gln Leu Leu Asp Ser Val
            -20                 -15                 -10
cta tgg ctg ggg gca cta gga ctg aca atc cag gca gtc ttt tcc acc   247
Leu Trp Leu Gly Ala Leu Gly Leu Thr Ile Gln Ala Val Phe Ser Thr
        -5                  1               5
act ggc cca gcc ctg ctg ctg ctt ctg gtc agc ttc ctc acc ttt gac   295
Thr Gly Pro Ala Leu Leu Leu Leu Leu Val Ser Phe Leu Thr Phe Asp
10                  15                  20                  25
ctg ctc cat agg ccc gca ggt cac act ctg cca cag cgc aaa ctt ctc   343
Leu Leu His Arg Pro Ala Gly His Thr Leu Pro Gln Arg Lys Leu Leu
                30                  35                  40
acc agg ggc cag agt cag ggg gcc ggt gaa ggt cct gga cag cag gag   391
Thr Arg Gly Gln Ser Gln Gly Ala Gly Glu Gly Pro Gly Gln Gln Glu
            45                  50                  55
gct cta ctc ctg caa atg ggt aca gtc tca gga caa ctt agc ctc cag   439
Ala Leu Leu Leu Gln Met Gly Thr Val Ser Gly Gln Leu Ser Leu Gln
        60                  65                  70
gac gca ctg ctg ctg ctg ctc atg ggg ctg ggc ccg ctc ctg aga gcc   487
Asp Ala Leu Leu Leu Leu Leu Met Gly Leu Gly Pro Leu Leu Arg Ala
    75                  80                  85
tgt ggc atg ccc ttg acc ctg ctt ggc ctg gct ttc tgc ctc cat cct   535
Cys Gly Met Pro Leu Thr Leu Leu Gly Leu Ala Phe Cys Leu His Pro
90                  95                  100                 105
tgg gcc tgagagcccc tccccacaac tcagtgtcct tcaaatatac aatgaccacc   591
Trp Ala
cttcttcaaa aaaa                                                    605
```

<210> 82
<211> 396
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 46..285

<221> sig_peptide
<222> 46..150
<223> Von Heijne matrix
      score 3.6
      seq LEPGLSSSAACNG/KE

<221> polyA_signal
<222> 364..369

<221> polyA_site
<222> 385..396

<400> 82
```
cctctacagg aatcagactc agcctctttt ggttttcagt gaagt atg cct ttt caa     57
                                                     Met Pro Phe Gln
                                                     -35
ttt gga acc cag cca agg agg ttt cca gtg gaa gga gga gat tct tca     105
Phe Gly Thr Gln Pro Arg Arg Phe Pro Val Glu Gly Gly Asp Ser Ser
    -30             -25             -20
att gag ctg gaa cct ggg ctg agc tcc agt gct gcc tgt aat ggg aag     153
Ile Glu Leu Glu Pro Gly Leu Ser Ser Ser Ala Ala Cys Asn Gly Lys
-15             -10             -5                  1
gag atg tca cca acc agg caa ctc cgg agg tgc cct gga agt cat tgc     201
Glu Met Ser Pro Thr Arg Gln Leu Arg Arg Cys Pro Gly Ser His Cys
            5               10              15
ctg aca ata act gat gtt ccc gtc act gtt tat gca aca acg aga aag     249
Leu Thr Ile Thr Asp Val Pro Val Thr Val Tyr Ala Thr Thr Arg Lys
        20              25              30
cca cct gca caa agc agc aag gaa atg cat cct aaa tagcaccatt          295
Pro Pro Ala Gln Ser Ser Lys Glu Met His Pro Lys
    35              40              45
aagtcttttg tcaaggtctg actaggtcaa gggtaatgga ccagtatcat ctggtgatct  355
ggtaaacaaa taaaagtggt ggcaccttca aaaaaaaaaa a                       396
```


<210> 83
<211> 432
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 22..240

<221> sig_peptide
<222> 22..84
<223> Von Heijne matrix
      score 12
      seq VLVLCVLLLQAQG/GY

<221> polyA_signal
<222> 397..402

<221> polyA_site
<222> 421..432

<400> 83
```
gctcacgctc tggtcagagt t atg gca ccc cag act ctg ctg cct gtc ctg     51
                        Met Ala Pro Gln Thr Leu Leu Pro Val Leu
                        -20             -15
gtt ctc tgt gtg ctg ctg ctg cag gcc cag gga gga tac cgt gac aag     99
Val Leu Cys Val Leu Leu Leu Gln Ala Gln Gly Gly Tyr Arg Asp Lys
-10             -5              1               5
atg agg atg cag aga atc aag gtc tgt gag aag cga ccc agc ata gat     147
Met Arg Met Gln Arg Ile Lys Val Cys Glu Lys Arg Pro Ser Ile Asp
            10              15              20
cta tgc atc cac cac tgt tca tgt ttc caa aag tgt gaa aca aat aag     195
Leu Cys Ile His His Cys Ser Cys Phe Gln Lys Cys Glu Thr Asn Lys
            25              30              35
ata tgc tgt tca gcc ttc tgt ggg aac att tgt atg agc atc cta         240
Ile Cys Cys Ser Ala Phe Cys Gly Asn Ile Cys Met Ser Ile Leu
            40              45              50
```

```
tgagtgggag agtgggctgg gatgtgcatc ctgctccctg aacccttcca tccgagactg    300
tgcccacatc cgaagcacaa ggacatcaaa tcatcagcac aagaacatca acaggaatgc    360
caccctcccc agtgtctgaa ctccctgtcc ctgtcaaatg aaccagaaca aatgcccatg    420
aaaaaaaaaa aa                                                        432
```


<210> 84
<211> 420
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 89..382

<221> polyA_site
<222> 408..420

<400> 84
```
gcttgcctga cccccatgtc gcctctgtag gtagaagaag tatgtcttcc tggacccct     60
ggctggtgct gtaacaaaga cccatgtg atg ctg ggg gca gag aca gag gag       112
                                 Met Leu Gly Ala Glu Thr Glu Glu
                                   1               5
aag ctg ttt gat gcc ccc ttg tcc atc agc aag aga gag cag ctg gaa      160
Lys Leu Phe Asp Ala Pro Leu Ser Ile Ser Lys Arg Glu Gln Leu Glu
     10              15                  20
cag cag gtc cca gag aac tac ttc tat gtg cca gac ctg ggc cag gtg      208
Gln Gln Val Pro Glu Asn Tyr Phe Tyr Val Pro Asp Leu Gly Gln Val
25                  30                  35                  40
cct gag att gat gtt cca tcc tac ctg cct gac ctg ccc ggc att gcc      256
Pro Glu Ile Asp Val Pro Ser Tyr Leu Pro Asp Leu Pro Gly Ile Ala
                45                  50                  55
aac gac ctc atg tac att gcc gac ctg ggc ccc ggc att gcc ccc tct      304
Asn Asp Leu Met Tyr Ile Ala Asp Leu Gly Pro Gly Ile Ala Pro Ser
                60                  65                  70
gcc cct ggc acc att cca gaa ctg ccc acc ttc cac act gag gta gcc      352
Ala Pro Gly Thr Ile Pro Glu Leu Pro Thr Phe His Thr Glu Val Ala
            75                  80                  85
gag cct ctc aag acc tac aag atg ggg tac taacagcacc accaccgccc        402
Glu Pro Leu Lys Thr Tyr Lys Met Gly Tyr
     90                  95
ccaccaaaaa aaaaaaaa                                                   420
```


<210> 85
<211> 501
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 80..415

<221> sig_peptide
<222> 80..142
<223> Von Heijne matrix
      score 5.4
      seq TFCLIFGLGAVWG/LG

<221> polyA_signal

<222> 471..476

<221> polyA_site
<222> 488..501

<400> 85
```
cccgcttgat tccaagaacc tcttcgatat ttattttttat ttttaaagag ggagacgatg     60
gactgagctg atccgcacc atg gag tct cgg gtc tta ctg aga aca ttc tgt    112
                     Met Glu Ser Arg Val Leu Leu Arg Thr Phe Cys
                         -20                     -15
ttg atc ttc ggt ctc gga gca gtt tgg ggg ctt ggt gtg gac cct tcc    160
Leu Ile Phe Gly Leu Gly Ala Val Trp Gly Leu Gly Val Asp Pro Ser
-10              -5                   1               5
cta cag att gac gtc tta aca gag tta gaa ctt ggg gag tcc acg acc    208
Leu Gln Ile Asp Val Leu Thr Glu Leu Glu Leu Gly Glu Ser Thr Thr
            10                  15                  20
gga gtg cgt cag gtc ccg ggg ctg cat aat ggg acg aaa gcc ttt ctc    256
Gly Val Arg Gln Val Pro Gly Leu His Asn Gly Thr Lys Ala Phe Leu
        25                  30                  35
ttt caa gat act ccc aga agc ata aaa gca tcc act gct aca gct gaa    304
Phe Gln Asp Thr Pro Arg Ser Ile Lys Ala Ser Thr Ala Thr Ala Glu
    40                  45                  50
cag ttt ttt cag aag ctg aga aat aaa cat gaa ttt act att ttg gtg    352
Gln Phe Phe Gln Lys Leu Arg Asn Lys His Glu Phe Thr Ile Leu Val
55                  60                  65                  70
acc cta aaa cag acc cac tta aat tca gga gtt att ctc tca att cac    400
Thr Leu Lys Gln Thr His Leu Asn Ser Gly Val Ile Leu Ser Ile His
            75                  80                  85
cac ttg gat cac agg taaatgtggt tgctggagtt tcctgtgttt tcattatatg    455
His Leu Asp His Arg
            90
tggttaaatg aatatattaa agagaagtaa acaaaaaaaa aaaaaa                   501
```


<210> 86
<211> 454
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 152..361

<221> sig_peptide
<222> 152..283
<223> Von Heijne matrix
        score 4.7
        seq FLLSLSLITYCFW/DP

<400> 86
```
gacattttac tttttttctgt taacgcttac cctagaaatt agaaatgaca ccacgtattc     60
ttagcgaagt ccagtttttca gcattttgtc cttattggac aatagcaagg atattagaac    120
gtgttggttc cgcgtgcttc cgtcttgagt t atg tgc tgc tat tgt cgg ata       172
                                  Met Cys Cys Tyr Cys Arg Ile
                                                          -40
ttt tgt ctt aga tgt acg tac ttt cct gtt cat tgt ggt atg tgt aat    220
Phe Cys Leu Arg Cys Thr Tyr Phe Pro Val His Cys Gly Met Cys Asn
        -35                 -30                 -25
ttg cgt tac ttt gaa ttt tcc acg ttt tta ctt tct ttg tct ctc atc    268
Leu Arg Tyr Phe Glu Phe Ser Thr Phe Leu Leu Ser Leu Ser Leu Ile
    -20                 -15                 -10
```

```
act tac tgc ttt tgg gac ccc ccc cat cgg ggt tca cat tcc ctc tcc    316
Thr Tyr Cys Phe Trp Asp Pro Pro His Arg Gly Ser His Ser Leu Ser
-5                    1              5              10
cta gag cac act ccc ttg gat ttc ctc gag tgg ggt ctg ctg cgg        361
Leu Glu His Thr Pro Leu Asp Phe Leu Glu Trp Gly Leu Leu Arg
            15              20              25
tgaagctttc ccattttatg tgcagattat tttcagaggg tatatagaat tcaggcagct  421
gtttcgttgt agcacattaa aaatattttc ccc                               454
```

<210> 87
<211> 1272
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 32..307

<221> sig_peptide
<222> 32..70
<223> Von Heijne matrix
      score 4.2
      seq MLFSLSLLSNLNQ/IG

<221> polyA_signal
<222> 1240..1245

<221> polyA_site
<222> 1261..1272

<400> 87
```
gtcaggttgc accgcccttt ggttcccgag c atg ctg ttt tct ctc agc ctt      52
                                    Met Leu Phe Ser Leu Ser Leu
                                                        -10
ctc tcc aac ctt aac caa atc ggc agc agc cac ctc gac cgc cca cac    100
Leu Ser Asn Leu Asn Gln Ile Gly Ser Ser His Leu Asp Arg Pro His
        -5                  1              5              10
att cct ggc caa tca gct cag ctg ttt att tac caa atg tct tca caa    148
Ile Pro Gly Gln Ser Ala Gln Leu Phe Ile Tyr Gln Met Ser Ser Gln
            15              20              25
caa cta cag cag cag cct tcg gct aac aaa aaa gca gga aaa atc cac    196
Gln Leu Gln Gln Gln Pro Ser Ala Asn Lys Lys Ala Gly Lys Ile His
            30              35              40
aac acc ccc ttc gcc aac caa cta aat cca acg caa cat ctg gca aaa    244
Asn Thr Pro Phe Ala Asn Gln Leu Asn Pro Thr Gln His Leu Ala Lys
            45              50              55
cct ttt cag caa att ctt cct ggc cgt cag tcc ggc agc ctc acc tca    292
Pro Phe Gln Gln Ile Leu Pro Gly Arg Gln Ser Gly Ser Leu Thr Ser
            60              65              70
cca ttt cta gct tgc tgaaacccaa aactaatctc caagaaggag aagcttctct    347
Pro Phe Leu Ala Cys
75
cgcagccgga gcaggtccct ttctagagat aggagaagag agagatcgct gtctcgggag  407
agaaatcaca agccgtcccg atccttctct aggtctcgta gtcgatttag gtcaaatgaa  467
aggaaataga agacagtttg caagagaagt ggtgtacagg aaattacttc atttgacagg  527
agtatgtaca gaaaattcaa gttttgtctg agacttcata agcttggtgc atttttaaga  587
tgttttagct gttcaaatct gtttgtctct tgaaacagtg acacaaaagt gtaattctct  647
atggtttgaa atggatcata cgaggcatgt aataccaaga attgttactt tacaatgttc  707
ccttaagcaa aattgaattt gctttgaact tttagttatg cacagactga taataaacct  767
ctaaacctgc ccagcggaag tgtgtttttt tttaaattta aatacagaaa caactggcaa  827
```

```
aaattgaact aagatttact ttttttttcca tagctgggat ataggctgca gctatagttg    887
aacaagcagt cttttaaaaac tgctgtgaaa cacaggccat cagggaaaac gaaatgctgc    947
actattaaat tagaggtttt tgaaaaatcc aactctcatc ctgggcagag gttgcctagt    1007
tggtatagaa tgttaagttt caagaaagtt tacctttgct ttaggtcgta agttccttat    1067
ttgattgccg tatatggata catggctgtt cgtgacattc tttatgtgca aatttgtgat    1127
ttcaaaaatg tcctgccagt ttaagggtac attgtagagc cgaactttga gttactgtgc    1187
aagatttttt ttcatgctgt catttgtaat atgttttgtg agaatccttg ggattaaagt    1247
tttggttaca gattaaaaaa aaaaa                                           1272
```

<210> 88
<211> 804
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 114..734

<221> sig_peptide
<222> 114..239
<223> Von Heijne matrix
      score 5.2
      seq LLFDLVCHEFCQS/DD

<221> polyA_signal
<222> 768..773

<221> polyA_site
<222> 793..804

<400> 88
```
ccaacaccag gaagagtctg aagagcagcc agtgtttcgg cttgtgccct gtatacttga    60
agctgccaaa caagtacggt agttctgaaa atccagaatg gcttgatgtt tac atg       116
                                                                     Met
cac att tta caa ctg ctt act aca gtg gat gat gga att caa gca att       164
His Ile Leu Gln Leu Leu Thr Thr Val Asp Asp Gly Ile Gln Ala Ile
    -40                 -35                 -30
gta cat tgt cct gac act gga aaa gac att tgg aat tta ctt ttt gac       212
Val His Cys Pro Asp Thr Gly Lys Asp Ile Trp Asn Leu Leu Phe Asp
-25                 -20                 -15                 -10
ctg gtc tgc cat gaa ttc tgc cag tct gat gat cca ccc atc att ctt       260
Leu Val Cys His Glu Phe Cys Gln Ser Asp Asp Pro Pro Ile Ile Leu
                -5                  1                   5
caa gaa cag aaa aca gtg cta gcc tct gtt ttt tca gtg ttg tct gcc       308
Gln Glu Gln Lys Thr Val Leu Ala Ser Val Phe Ser Val Leu Ser Ala
                10                  15                  20
atc tat gcc tca cag act gag caa gag tat cta aag ata gaa aaa gta       356
Ile Tyr Ala Ser Gln Thr Glu Gln Glu Tyr Leu Lys Ile Glu Lys Val
            25                  30                  35
gat ctt cct cta att gac agc ctc att cgg gtc tta caa aat atg gaa       404
Asp Leu Pro Leu Ile Asp Ser Leu Ile Arg Val Leu Gln Asn Met Glu
40                  45                  50                  55
cag tgt cag aaa aaa cca gag aac tcg gca gag tct aac aca gag gaa       452
Gln Cys Gln Lys Lys Pro Glu Asn Ser Ala Glu Ser Asn Thr Glu Glu
                60                  65                  70
act aaa agg act gat tta acc caa gat gat ctc cac ttg aaa atc tta       500
Thr Lys Arg Thr Asp Leu Thr Gln Asp Asp Leu His Leu Lys Ile Leu
                75                  80                  85
aag gat att tta tgt gaa ttt ctt tct aat att ttt cag gca tta aca       548
Lys Asp Ile Leu Cys Glu Phe Leu Ser Asn Ile Phe Gln Ala Leu Thr
```

```
           90                    95                   100
aag gag acg gtg gct cag gga gta aag gaa ggc cag ttg agc aaa cag    596
Lys Glu Thr Val Ala Gln Gly Val Lys Glu Gly Gln Leu Ser Lys Gln
    105                   110                  115
aag tgt tcc tct gca ttt caa aac ctt ctt cct ttc tat agc cct gtg    644
Lys Cys Ser Ser Ala Phe Gln Asn Leu Leu Pro Phe Tyr Ser Pro Val
120                   125                   130                  135
gtg gaa gat ttt att aaa atc cta cgt gaa gtt gat aag gcg ctt gct    692
Val Glu Asp Phe Ile Lys Ile Leu Arg Glu Val Asp Lys Ala Leu Ala
                    140                   145                  150
gat gac ttg gaa aaa aac ttc cca agt ttg aag gtt cag act           734
Asp Asp Leu Glu Lys Asn Phe Pro Ser Leu Lys Val Gln Thr
                155                   160                  165
taaaacctga attggaatta cttctgtaca agaaataaac tttattttc tcactgacaa    794
aaaaaaaaaa                                                         804
```

```
<210> 89
<211> 802
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 199..801

<221> polyA_signal
<222> 780..785

<221> polyA_site
<222> 791..802

<400> 89
agtcaccgcc tgcttcgcac tgagcctccc gactcagact ctgagtccag ctccgaagag    60
gaagaggaat tcggtgtggt tggaaatcgc tctcgctttg ccaagggaga ctatttacga   120
tgctgcaaga tctgttatcc gctctgtggt tttgtcatcc ttgctgcctg tgttgtggcc   180
tgtgttggct tggtgtgg atg cag gtt gct ctc aag gag gat ctg gat gcc    231
                     Met Gln Val Ala Leu Lys Glu Asp Leu Asp Ala
                      1               5                  10
ctc aag gaa aaa ttt cga aca atg gaa tct aat cag aaa agc tca ttc    279
Leu Lys Glu Lys Phe Arg Thr Met Glu Ser Asn Gln Lys Ser Ser Phe
                15                  20                  25
caa gaa atc ccc aaa ctt aat gaa gaa cta ctc agc aag caa aaa caa    327
Gln Glu Ile Pro Lys Leu Asn Glu Glu Leu Leu Ser Lys Gln Lys Gln
            30                  35                  40
ctt gag aag att gaa tct gga gag atg ggt ttg aac aaa gtc tgg ata    375
Leu Glu Lys Ile Glu Ser Gly Glu Met Gly Leu Asn Lys Val Trp Ile
        45                  50                  55
aac atc aca gaa atg aat aag cag att tct ctg ttg act tct gca gtg    423
Asn Ile Thr Glu Met Asn Lys Gln Ile Ser Leu Leu Thr Ser Ala Val
60                  65                  70                  75
aac cac ctc aaa gcc aat gtt aag tca gct gca gac ttg att agc ctg    471
Asn His Leu Lys Ala Asn Val Lys Ser Ala Ala Asp Leu Ile Ser Leu
                80                  85                  90
cct acc act gta gag gga ctt cag aag agt gta gct tcc att ggc aat    519
Pro Thr Thr Val Glu Gly Leu Gln Lys Ser Val Ala Ser Ile Gly Asn
                95                  100                 105
act tta aac agc gtc cat ctt gct gtg gaa gca cta cag aaa act gtg    567
Thr Leu Asn Ser Val His Leu Ala Val Glu Ala Leu Gln Lys Thr Val
        110                 115                 120
gat gaa cac aag aaa acg atg gaa tta ctg cag agt gat atg aat cag    615
```

```
            Asp Glu His Lys Lys Thr Met Glu Leu Leu Gln Ser Asp Met Asn Gln
                125                 130                 135
cac ttc ttg aag gag act cct gga agc aac cag atc att ccg tca cct        663
His Phe Leu Lys Glu Thr Pro Gly Ser Asn Gln Ile Ile Pro Ser Pro
140             145                 150                 155
tca gcc aca tca gaa ctt gac aat aaa acc cac agt gag aat ttg aaa        711
Ser Ala Thr Ser Glu Leu Asp Asn Lys Thr His Ser Glu Asn Leu Lys
                160                 165                 170
cag atg ggt gat aga tct gcc act ctg aaa aga cag tct ttg gac caa        759
Gln Met Gly Asp Arg Ser Ala Thr Leu Lys Arg Gln Ser Leu Asp Gln
                175             180                 185
gtc acc aac aga aca gat aca gta aaa atc caa aaa aaa aaa a             802
Val Thr Asn Arg Thr Asp Thr Val Lys Ile Gln Lys Lys Lys
            190                 195                 200
```

```
<210> 90
<211> 1490
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 38..1174

<221> sig_peptide
<222> 38..148
<223> Von Heijne matrix
      score 7.3
      seq LLSACLVTLWGLG/EP

<221> polyA_signal
<222> 1452..1457

<221> polyA_site
<222> 1478..1490

<400> 90
tcatcatcca gagcagccag tgtccgggag gcagaag atg ccc cac tcc agc ctg        55
                                         Met Pro His Ser Ser Leu
                                                             -35
cat cca tcc atc ccg tgt ccc agg ggt cac ggg gcc cag aag gca gcc        103
His Pro Ser Ile Pro Cys Pro Arg Gly His Gly Ala Gln Lys Ala Ala
    -30             -25                 -20
ttg gtt ctg ctg agt gcc tgc ctg gtg acc ctt tgg ggg cta gga gag        151
Leu Val Leu Leu Ser Ala Cys Leu Val Thr Leu Trp Gly Leu Gly Glu
-15                 -10                 -5                  1
cca cca gag cac act ctc cgg tac ctg gtc ctc cac cta gcc tcc ctg        199
Pro Pro Glu His Thr Leu Arg Tyr Leu Val Leu His Leu Ala Ser Leu
            5                   10                  15
cag ctg gga ctg ctg tta aac ggg gtc tgc agc ctg gct gag gag ctg        247
Gln Leu Gly Leu Leu Leu Asn Gly Val Cys Ser Leu Ala Glu Glu Leu
        20                  25                  30
cgc cac atc cac tcc agg tac cgg ggc agc tac tgg agg act gtg cgg        295
Arg His Ile His Ser Arg Tyr Arg Gly Ser Tyr Trp Arg Thr Val Arg
        35                  40                  45
gcc tgc ctg ggc tgc ccc ctc cgc cgt ggg gcc ctg ttg ctg ctg tcc        343
Ala Cys Leu Gly Cys Pro Leu Arg Arg Gly Ala Leu Leu Leu Leu Ser
50                  55                  60                  65
atc tat ttc tac tac tcc ctc cca aat gcg gtc ggc ccg ccc ttc act        391
Ile Tyr Phe Tyr Tyr Ser Leu Pro Asn Ala Val Gly Pro Pro Phe Thr
```

```
                      70                   75                   80
tgg atg ctt gcc ctc ctg ggc ctc tcg cag gca ctg aac atc ctc ctg      439
Trp Met Leu Ala Leu Leu Gly Leu Ser Gln Ala Leu Asn Ile Leu Leu
                 85                   90                   95
ggc ctc aag ggc ctg gcc cca gct gag atc tct gca gtg tgt gaa aaa      487
Gly Leu Lys Gly Leu Ala Pro Ala Glu Ile Ser Ala Val Cys Glu Lys
           100                  105                  110
ggg aat ttc aac gtg gcc cat ggg ctg gca tgg tca tat tac atc gga      535
Gly Asn Phe Asn Val Ala His Gly Leu Ala Trp Ser Tyr Tyr Ile Gly
           115                  120                  125
tat ctg cgg ctg atc ctg cca gag ctc cag gcc cgg att cga act tac      583
Tyr Leu Arg Leu Ile Leu Pro Glu Leu Gln Ala Arg Ile Arg Thr Tyr
130                  135                  140                  145
aat cag cat tac aac aac ctg cta cgg ggt gca gtg agc cag cgg ctg      631
Asn Gln His Tyr Asn Asn Leu Leu Arg Gly Ala Val Ser Gln Arg Leu
                 150                  155                  160
tat att ctc ctc cca ttg gac tgt ggg gtg cct gat aac ctg agt atg      679
Tyr Ile Leu Leu Pro Leu Asp Cys Gly Val Pro Asp Asn Leu Ser Met
                 165                  170                  175
gct gac ccc aac att cgc ttc ctg gat aaa ctg ccc cag cag acc ggt      727
Ala Asp Pro Asn Ile Arg Phe Leu Asp Lys Leu Pro Gln Gln Thr Gly
           180                  185                  190
gac cgt gct ggc atc aag gat cgg gtt tac agc aac agc atc tat gag      775
Asp Arg Ala Gly Ile Lys Asp Arg Val Tyr Ser Asn Ser Ile Tyr Glu
           195                  200                  205
ctt ctg gag aac ggg cag cgg gcg ggc acc tgt gtc ctg gag tac gcc      823
Leu Leu Glu Asn Gly Gln Arg Ala Gly Thr Cys Val Leu Glu Tyr Ala
210                  215                  220                  225
acc ccc ttg cag act ttg ttt gcc atg tca caa tac agt caa gct ggc      871
Thr Pro Leu Gln Thr Leu Phe Ala Met Ser Gln Tyr Ser Gln Ala Gly
                 230                  235                  240
ttt agc cgg gag gat agg ctt gag cag gcc aaa ctc ttc tgc cgg aca      919
Phe Ser Arg Glu Asp Arg Leu Glu Gln Ala Lys Leu Phe Cys Arg Thr
                 245                  250                  255
ctt gag gac atc ctg gca gat gcc cct gag tct cag aac aac tgc cgc      967
Leu Glu Asp Ile Leu Ala Asp Ala Pro Glu Ser Gln Asn Asn Cys Arg
           260                  265                  270
ctc att gcc tac cag gaa cct gca gat gac agc agc ttc tcg ctg tcc      1015
Leu Ile Ala Tyr Gln Glu Pro Ala Asp Asp Ser Ser Phe Ser Leu Ser
           275                  280                  285
cag gag gtt ctc cgg cac ctg cgg cag gag gaa aag gaa gag gtt acc      1063
Gln Glu Val Leu Arg His Leu Arg Gln Glu Glu Lys Glu Glu Val Thr
290                  295                  300                  305
gtg ggc agc ttg aag acc tca gcg gtg ccc agt acc tcc acg atg tcc      1111
Val Gly Ser Leu Lys Thr Ser Ala Val Pro Ser Thr Ser Thr Met Ser
                 310                  315                  320
caa gag cct gag ctc ctc ctc agt gga atg gga aag ccc ctc cct ctc      1159
Gln Glu Pro Glu Leu Leu Leu Ser Gly Met Gly Lys Pro Leu Pro Leu
                 325                  330                  335
cgc acg gat ttc tct tgagacccag ggtcaccagg ccagagcctc cagtggtctc      1214
Arg Thr Asp Phe Ser
                 340
caagcctctg gactggggggc tctcttcagt ggctgaatgt ccagcagagc tatttccttc    1274
cacagggggc cttgcaggga agggtccagg acttgacatc ttaagatgcg tcttgtcccc    1334
ttgggccagt catttcccct ctctgagcct cggtgtcttc aacctgtgaa atgggatcat    1394
aatcactgcc ttacctccct .cacggttgtt gtgaggactg agtgtgtgga agttttcat    1454
aaactttgga tgctagtgta cttaaaaaaa aaaaaa                              1490
```

```
<210> 91
<211> 361
```

```
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 26..361

<221> polyA_site
<222> 350..361

<400> 91
tcgagaagct gccccttagc caacc atg ccg tct gag ggt cgc tgc tgg gag        52
                            Met Pro Ser Glu Gly Arg Cys Trp Glu
                             1               5

acc ttg aag gcc cta cgc agt tcc gac aaa ggt cgc ctt tgc tac tac       100
Thr Leu Lys Ala Leu Arg Ser Ser Asp Lys Gly Arg Leu Cys Tyr Tyr
10                 15                  20                  25

cgc gac tgg ctg ctg cgg cgc gag gat gtt tta gaa gaa tgt atg tct       148
Arg Asp Trp Leu Leu Arg Arg Glu Asp Val Leu Glu Glu Cys Met Ser
                30                  35                  40

ctt ccc aag cta tct tct tat tct gga tgg gtg gta gag cac gtc cta       196
Leu Pro Lys Leu Ser Ser Tyr Ser Gly Trp Val Val Glu His Val Leu
            45                  50                  55

ccc cat atg cag gag aac caa cct ctg tct gag act tcg cca tcc tct       244
Pro His Met Gln Glu Asn Gln Pro Leu Ser Glu Thr Ser Pro Ser Ser
        60                  65                  70

acg tca gct tca gcc cta gat caa ccc tca ttt gtt ccc aaa tct cct       292
Thr Ser Ala Ser Ala Leu Asp Gln Pro Ser Phe Val Pro Lys Ser Pro
    75                  80                  85

gac gca agc tct gcc ttt tcc cca gcc tcc cct gca aca cca aat gga       340
Asp Ala Ser Ser Ala Phe Ser Pro Ala Ser Pro Ala Thr Pro Asn Gly
90                  95                 100                 105

acc aag ggc aaa aaa aaa aaa                                           361
Thr Lys Gly Lys Lys Lys Lys
                110


<210> 92
<211> 605
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 3..131

<221> polyA_site
<222> 591..605

<400> 92
ca tcc ctt ccc cag gct tta tgg ttc cag ttc ttc tac cac tct gga        47
   Ser Leu Pro Gln Ala Leu Trp Phe Gln Phe Phe Tyr His Ser Gly
    1           5                  10                  15

agc tcc cta gaa tct cct gga atg ctt aat gga cct ttc cag cac cga        95
Ser Ser Leu Glu Ser Pro Gly Met Leu Asn Gly Pro Phe Gln His Arg
            20                  25                  30

aat tca aga att atg act cat cgg tca gca gaa aag tgaggatacc           141
Asn Ser Arg Ile Met Thr His Arg Ser Ala Glu Lys
            35                  40

tttcctaac ctacctgctt cccctgcagt ttcctcacaa tcttactctt tatattttag     201
catatgtagc ttctcaggat gttaattctg ttctctctgt gttggtgtct gagcacccag    261
```

```
aaggtagagc caggggcact tataaaccag gagcattatt tgacaggcac ttaagaaaga    321
cactggctac gtaatcccag cactttggga ggctgaggcg gatggatcac atgaggtcag    381
gagttcgaga ccagcctggc cagcatggtg aaaccctgtc tctactaaaa atacaaaaat    441
tagctgggtg tggttgcaca cgcctgtaat cccagctacc tgggaggctg aggcaggaga    501
atcgcttgaa cttgggaggc ggaggttgca gtgagcctag attttgccat tgcactccag    561
cctgggtgac aagggcgaaa ctccatccca aaaaaaaaaa aaaa                     605
```

<210> 93
<211> 591
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 33..185

<221> sig_peptide
<222> 33..80
<223> Von Heijne matrix
      score 3.7
      seq IALTLIPSMLSRA/AG

<221> polyA_signal
<222> 570..575

<221> polyA_site
<222> 586..591

<400> 93
```
caatcttctc agcttataac cgtctttccc tt atg cta agg ata gcc ctt aca     53
                                     Met Leu Arg Ile Ala Leu Thr
                                         -15                 -10
ctc atc cca tct atg ctg tca agg gct gct ggt tgg tgc tgg tac aag    101
Leu Ile Pro Ser Met Leu Ser Arg Ala Ala Gly Trp Cys Trp Tyr Lys
              -5                  1               5
gag ccc act cag cag ttt tct tac ctt tgc ctg ccc tgc ctt tca tgg    149
Glu Pro Thr Gln Gln Phe Ser Tyr Leu Cys Leu Pro Cys Leu Ser Trp
        10                  15                  20
aat aag aaa ggc aac gtt ttg cag ctt cca aat ttc tgaagaaact         195
Asn Lys Lys Gly Asn Val Leu Gln Leu Pro Asn Phe
        25                  30                  35
aatctcagat tggcagttaa agtcaaaatg ttgccaaata tttattcctt ttgcctaagt    255
ttggctaccc ggttcaattg ctttttattt ttaatgtctt gactcttcag agttcgtacc    315
tcaaaagaac aatgagaaca tttgctttgc tttctgctga atccctaatc tcaacaatct    375
atacctggac tgtccagttc tcctcctgtg ctatcttctc ttctatccaa gtagaatgta    435
tgccaggagc tccttccctc tagcaatttc tactaaaatg tccaagtaga atgtttcctt    495
ttacaatcaa attactgtat ttattaattt gctagaatcc agtaaatcat tttggtagct    555
ctggctgtgc tatcaataaa aagatgaaag caaaaa                              591
```

<210> 94
<211> 1150
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 184..915

<221> sig_peptide
<222> 184..237
<223> Von Heijne matrix
     score 3.5
     seq LLGLELSEAEAIG/AD

<221> polyA_signal
<222> 1119..1124

<221> polyA_site
<222> 1139..1150

<400> 94
```
cggatttgac datggtgttc ggtcttgaat ggaaatgtag tcttaggcca gtcttaggtt          60
tttgaacagg atagtaggta tccgggagtcg attgagggcc agagcaggca ctggggttcg         120
gatcctgggc aaagtttccc acgttgaggg tctcgaggac gcctagatct ctttcccagg         180
gcc atg gcg aac ccg aag ctg ctg gga ctg gag cta agc gag gcg gag          228
    Met Ala Asn Pro Lys Leu Leu Gly Leu Glu Leu Ser Glu Ala Glu
                -15                 -10                 -5
gcg atc ggt gct gat tcg gcg cga ttt gag gag ctg ctg ctg cag gcc          276
Ala Ile Gly Ala Asp Ser Ala Arg Phe Glu Glu Leu Leu Leu Gln Ala
        1               5                   10
tcg aag gag ctc cag caa gcc cag aca acc aga cca gaa tcg aca caa          324
Ser Lys Glu Leu Gln Gln Ala Gln Thr Thr Arg Pro Glu Ser Thr Gln
    15                  20                  25
atc cag cct cag cct ggt ttc tgc ata aag acc aac tcc tcg gaa ggg          372
Ile Gln Pro Gln Pro Gly Phe Cys Ile Lys Thr Asn Ser Ser Glu Gly
30                  35                  40                  45
aag gtt ttc atc aac atc tgc cac tcc ccc tct atc cct cct ccc gcc          420
Lys Val Phe Ile Asn Ile Cys His Ser Pro Ser Ile Pro Pro Pro Ala
                50                  55                  60
gac gtg acc gag gag gag ctg ctt cag atg cta gag gag gac caa gct          468
Asp Val Thr Glu Glu Glu Leu Leu Gln Met Leu Glu Glu Asp Gln Ala
            65                  70                  75
ggg ttt cgc atc ccc atg agt ctg gga gag cct cat gca gaa ctg gat          516
Gly Phe Arg Ile Pro Met Ser Leu Gly Glu Pro His Ala Glu Leu Asp
        80                  85                  90
gca aaa ggc cag gga tgt acc gcc tac gac gta gct gtc aac agc gac          564
Ala Lys Gly Gln Gly Cys Thr Ala Tyr Asp Val Ala Val Asn Ser Asp
    95                  100                 105
ttc tac cgg agg atg cag aac agc gat ttc ttg cgg gag ctc gtg atc          612
Phe Tyr Arg Arg Met Gln Asn Ser Asp Phe Leu Arg Glu Leu Val Ile
110                 115                 120                 125
acc atc gcc agg gag ggc ctt gag gac ata tac aac ttg cag ctg aat          660
Thr Ile Ala Arg Glu Gly Leu Glu Asp Ile Tyr Asn Leu Gln Leu Asn
                130                 135                 140
ccg gaa tgg cgc atg atg aag aac cgg cca ttc atg ggc tcc atc tcg          708
Pro Glu Trp Arg Met Met Lys Asn Arg Pro Phe Met Gly Ser Ile Ser
            145                 150                 155
cag cag aac atc cgc tcg gag cag cgt cct cgg atc cag gag ctg ggg          756
Gln Gln Asn Ile Arg Ser Glu Gln Arg Pro Arg Ile Gln Glu Leu Gly
            160                 165                 170
gac ctg tac acg ccc gcc ccc ggg aga gct gag tca ggg cct gaa aag          804
Asp Leu Tyr Thr Pro Ala Pro Gly Arg Ala Glu Ser Gly Pro Glu Lys
    175                 180                 185
cct cac ctg aac ctg tgg ctg gaa gcc ccc gac ctc ctc ttg gcc gaa          852
Pro His Leu Asn Leu Trp Leu Glu Ala Pro Asp Leu Leu Leu Ala Glu
190                 195                 200                 205
gtt gac ctc ccc aaa ctg gat gga gcc ctg ggg ctg tcg ctg gag atc          900
Val Asp Leu Pro Lys Leu Asp Gly Ala Leu Gly Leu Ser Leu Glu Ile
                210                 215                 220
ggg aga acc gcc tgg tgatggggggg cccccagcag ctgtatcatc tagacgctta         955
Gly Arg Thr Ala Trp
```

```
Gly Arg Thr Ala Trp
        225
tatcccgccg cagatcaact ctcatgagag caaggcagcc ttccaccgga agagaaagca   1015
attaatggtg gccatgccgc ttctgccggt gccttcttga tcagggtgtc tccttgtgct   1075
tctgagatgt ggagaagagg ctgctggctt ccctaaaagt tgaaataaaa gattttttgcc  1135
tttaaaaaaa aaaaa                                                     1150
```

```
<210> 95
<211> 1513
<212> DNA
<213> Homo sapiens

<220>

<221> CDS
<222> 58..1115

<221> sig_peptide
<222> 58..159
<223> Von Heijne matrix
      score 4
      seq IAVLYLHLYDVFG/DP

<221> polyA_signal
<222> 1486..1491

<221> polyA_site
<222> 1504..1513


<400> 95
ctgactcctg agttctcaca acgcttgacc aataagattc gggagcttct tcagcaa           57
atg gag aga ggc ctg aaa tca gca gac cct cgg gat ggc acc ggt tac       105
Met Glu Arg Gly Leu Lys Ser Ala Asp Pro Arg Asp Gly Thr Gly Tyr
            -30                 -25                 -20
act ggc tgg gca ggt att gct gtg ctt tac tta cat ctt tat gat gta       153
Thr Gly Trp Ala Gly Ile Ala Val Leu Tyr Leu His Leu Tyr Asp Val
            -15                 -10                 -5
ttt ggg gac cct gcc tac cta cag tta gca cat ggc tat gta aag caa       201
Phe Gly Asp Pro Ala Tyr Leu Gln Leu Ala His Gly Tyr Val Lys Gln
        1               5                   10
agt ctg aac tgc tta acc aag cgc tcc atc acc ttc ctt tgt ggg gat       249
Ser Leu Asn Cys Leu Thr Lys Arg Ser Ile Thr Phe Leu Cys Gly Asp
    15                  20                  25                  30
gca ggc ccc ctg gca gtg gcc gct gtg cta tat cat aag atg aac aat       297
Ala Gly Pro Leu Ala Val Ala Ala Val Leu Tyr His Lys Met Asn Asn
                35                  40                  45
gag aag cag gca gaa gat tgc atc aca cgg cta att cac cta aat aag       345
Glu Lys Gln Ala Glu Asp Cys Ile Thr Arg Leu Ile His Leu Asn Lys
            50                  55                  60
att gat cct cat gct cca aat gaa atg ctc tat ggg cga ata ggc tac       393
Ile Asp Pro His Ala Pro Asn Glu Met Leu Tyr Gly Arg Ile Gly Tyr
            65                  70                  75
atc tat gct ctt ctt ttt gtc aat aag aac ttt gga gtg gaa aag act       441
Ile Tyr Ala Leu Leu Phe Val Asn Lys Asn Phe Gly Val Glu Lys Thr
        80                  85                  90
cct caa agc cat att cag cag att tgt gaa aca att tta acc tct gga       489
Pro Gln Ser His Ile Gln Gln Ile Cys Glu Thr Ile Leu Thr Ser Gly
    95                  100                 105                 110
gaa aac cta gct agg aag aga aac ttc acg gca aag tct cca ctg atg       537
Glu Asn Leu Ala Arg Lys Arg Asn Phe Thr Ala Lys Ser Pro Leu Met
                115                 120                 125
```

```
tat gaa tgg tac cag gaa tat tat gta ggg gct gct cat ggc ctg gct      585
Tyr Glu Trp Tyr Gln Glu Tyr Tyr Val Gly Ala Ala His Gly Leu Ala
            130                 135                 140
gga att tat tac tac ctg atg cag ccc agc ctt caa gtg agc caa ggg      633
Gly Ile Tyr Tyr Tyr Leu Met Gln Pro Ser Leu Gln Val Ser Gln Gly
            145                 150                 155
aag tta cat agt ttg gtc aag ccc agt gta gac tac gtc tgc cag ctg      681
Lys Leu His Ser Leu Val Lys Pro Ser Val Asp Tyr Val Cys Gln Leu
    160                 165                 170
aaa ttc cct tct ggc aat tac cct cca tgt ata ggt gat aat cga gat      729
Lys Phe Pro Ser Gly Asn Tyr Pro Pro Cys Ile Gly Asp Asn Arg Asp
175                 180                 185                 190
ctg ctt gtc cat tgg tgc cat ggc gcc cct ggg gta atc tac atg ctc      777
Leu Leu Val His Trp Cys His Gly Ala Pro Gly Val Ile Tyr Met Leu
                195                 200                 205
atc cag gcc tat aag gta ttc aga gag gaa aag tat ctc tgt gat gcc      825
Ile Gln Ala Tyr Lys Val Phe Arg Glu Glu Lys Tyr Leu Cys Asp Ala
            210                 215                 220
tat cag tgt gct gat gtg atc tgg caa tat ggg ttg ctg aag aag gga      873
Tyr Gln Cys Ala Asp Val Ile Trp Gln Tyr Gly Leu Leu Lys Lys Gly
            225                 230                 235
tat ggg ctg tgc cac ggt tct gca ggg aat gcc tat gcc ttc ctg aca      921
Tyr Gly Leu Cys His Gly Ser Ala Gly Asn Ala Tyr Ala Phe Leu Thr
            240                 245                 250
ctc tac aac ctc aca cag gac atg aag tac ctg tat agg gcc tgt aag      969
Leu Tyr Asn Leu Thr Gln Asp Met Lys Tyr Leu Tyr Arg Ala Cys Lys
255                 260                 265                 270
ttt gct gaa tgg tgc tta gag tat gga gaa cat gga tgc aga aca cca     1017
Phe Ala Glu Trp Cys Leu Glu Tyr Gly Glu His Gly Cys Arg Thr Pro
                275                 280                 285
gac acc cct ttc tct ctc ttt gaa gga atg gct ggg aca ata tat ttc     1065
Asp Thr Pro Phe Ser Leu Phe Glu Gly Met Ala Gly Thr Ile Tyr Phe
            290                 295                 300
ctg gct gac ctg cta gtc ccc aca aaa gcc agg ttc cct gca ttt gaa     1113
Leu Ala Asp Leu Leu Val Pro Thr Lys Ala Arg Phe Pro Ala Phe Glu
            305                 310                 315
ctc tgaaaggata gcatgccacc tgcaactcac tgcatgaccc tttctgtata          1166
Leu
ttcaaaccca agctaagtgc ttccgttgct ttccaaggaa acaaagagtc aaactgtgga    1226
cttgattttg ttagcttttt tcagaattta tctttcattc agttcccttc cattatcatt   1286
tactttact tagaagtatc caaggaagtc ttttaacttt aatttccatt tcttcctaaa    1346
gggagagtga gtgatatgta cagtgttttg agattgtata catatattcc agaacttgga   1406
ggaaatctta tttaagttta tgaatataac catctgttac tgttctaaaa atgtttaaaa   1466
gaaactcaat acagataaag ataaatatgt gactattaaa aaaaaaa                 1513
```

```
<210> 96
<211> 417
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 327..416

<221> polyA_site
<222> 404..417

<400> 96
tgttttgagg tgttggcatt cttcgctgat ttggctgttc ccaatgttta cattatttaa    60
tcttgcaaaa atggttctgt gcacttggat gtgaaatgct ttccagtttt atttttttta   120
```

```
tgttgttatc cttggatgta caaaaaattc agaaaatgat ctctgtagat attctgtttt      180
attttggtca tctttagaag ttatcaggaa tgtgtttaaa acaagaagag aactttttcta    240
aggaatgata catagaaaag attttatttt aaaatgagtt gtaaagcttg tgtttctttg     300
ttgctgcaag ctatctgccc aagtta atg caa atg gac aca ttt ttt atg tca     353
                                Met Gln Met Asp Thr Phe Phe Met Ser
                                 1                   5
gaa aaa cac aca cac aca cac aca cat ata cac aca cac aca cga aaa      401
Glu Lys His Thr His Thr His Thr His Ile His Thr His Thr Arg Lys
10              15                  20                  25
aca aaa aaa aaa aaa a                                                 417
Thr Lys Lys Lys Lys
            30
```

<210> 97
<211> 603
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 63..398

<221> sig_peptide
<222> 63..206
<223> Von Heijne matrix
      score 4.9
      seq PSLAAGLLFGSLA/GL

<400> 97
```
ggggccttcg tgagaccggt gcaggcctgg ggtagtctcc tgtctggaca gagaagagaa      60
aa atg cag gac act ggc tca gta gtg cct ttg cat tgg ttt ggc ttt       107
   Met Gln Asp Thr Gly Ser Val Val Pro Leu His Trp Phe Gly Phe
           -45                 -40                 -35
ggc tac gca gca ctg gtt gct tct ggt ggg atc att ggc tat gta aaa     155
Gly Tyr Ala Ala Leu Val Ala Ser Gly Gly Ile Ile Gly Tyr Val Lys
        -30                 -25                 -20
gca ggc agc gtg ccg tcc ctg gct gca ggg ctg ctc ttt ggc agt cta     203
Ala Gly Ser Val Pro Ser Leu Ala Ala Gly Leu Leu Phe Gly Ser Leu
        -15                 -10                 -5
gcc ggc ctg ggt gct tac cag ctg tct cag gat cca agg aac gtt tgg     251
Ala Gly Leu Gly Ala Tyr Gln Leu Ser Gln Asp Pro Arg Asn Val Trp
 1               5                   10                  15
gtt ttc cta gct aca tct ggt acc ttg gct ggc att atg gga atg agg     299
Val Phe Leu Ala Thr Ser Gly Thr Leu Ala Gly Ile Met Gly Met Arg
        20                  25                  30
ttc tac cac tct gga aaa ttc atg cct gca ggt tta att gca ggt gcc     347
Phe Tyr His Ser Gly Lys Phe Met Pro Ala Gly Leu Ile Ala Gly Ala
        35                  40                  45
agt ttg ctg atg gtc gcc aaa gtt gga gtt agt atg ttc aac aga ccc     395
Ser Leu Leu Met Val Ala Lys Val Gly Val Ser Met Phe Asn Arg Pro
        50                  55                  60
cat tagcagaagt catgttccag cttagactga tgaagaatta aaaatctgca           448
His
tcttccacta tttttcaatat attaagagaa ataagtgcag cattttttgca tctgacattt   508
tacctaaaaa aaaagacacc aaacttggca gagaggtgga aaatcagtca tgattacaaa    568
cctacagagg tggcgagtat gtaacacaag agctt                               603
```

<210> 98

```
<211> 522
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 2..163

<221> polyA_signal
<222> 488..493

<221> polyA_site
<222> 511..522

<400> 98
c gag att gcg ggc tat ggc gcc gaa ggt ttt tcg tca gta ctg gga tat      49
  Glu Ile Ala Gly Tyr Gly Ala Glu Gly Phe Ser Ser Val Leu Gly Tyr
  1              5                  10                  15
ccc cga tgg cac cga ttg cca ccg caa agc cta cag cac cac cag tat        97
Pro Arg Trp His Arg Leu Pro Pro Gln Ser Leu Gln His His Gln Tyr
         20                  25                  30
tgc cag cgt cgc tgg cct gac cgc cgc tgc cta cag agt cac act caa       145
Cys Gln Arg Arg Trp Pro Asp Arg Arg Cys Leu Gln Ser His Thr Gln
        35                  40                  45
tcc tcc ggg cac ctt cct nntgaaggag tggctaaggt tggacaatac             193
Ser Ser Gly His Leu Pro
   50
acgttcactg cagctgctgt cggggccgtg tttggcctca ccacctgcat cagcgcccat    253
gtccgcgaga agcccgacga ccccctgaac tacttccccg gtggctgcgc cnggaggcct    313
gactctggga gcacgcacgc acaactacgg gattggcgcc gccgcctgcg tgtactttgg    373
catagcggcc tccctggtca agatgggccg gctggagggc tgggaggtgt ttgcaaaacc    433
caaggtgtga gccctgtgcc tgccgggacc tccagcctgc agaatgcgtc cagaaataaa    493
ttctgtgtct gtgtgtgaaa aaaaaaaaa                                     522


<210> 99
<211> 956
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 13..465

<221> sig_peptide
<222> 13..75
<223> Von Heijne matrix
      score 3.9
      seq PVAVTAAVAPVLS/IN

<400> 99
ngagtcggga aa atg gct gcg agt acn tcn atg gnc ccg gtg gct gtg acg    51
              Met Ala Ala Ser Thr Ser Met Xaa Pro Val Ala Val Thr
                 -20                 -15                 -10
gcg gca gtg gcg cct gtc ctg tcc ata aac agc gat ttc tca gat ttg      99
Ala Ala Val Ala Pro Val Leu Ser Ile Asn Ser Asp Phe Ser Asp Leu
         -5                  1                   5
cgg gaa att aaa aag caa ctg ctg ctt att gcg ggc ctt acc cgg gag     147
Arg Glu Ile Lys Lys Gln Leu Leu Leu Ile Ala Gly Leu Thr Arg Glu
   10                  15                  20
cgg ggc cta cta cac agt agc aaa tgg tcg gcg gag ttg gct ttc tct     195
```

```
Arg Gly Leu Leu His Ser Ser Lys Trp Ser Ala Glu Leu Ala Phe Ser
25              30              35              40
ctc cct gca ttg cct cnt ggc cag ctg caa ccg cct ccg cct att aca      243
Leu Pro Ala Leu Pro Xaa Gly Gln Leu Gln Pro Pro Pro Ile Thr
                45              50              55
gag gaa gat gcc cag gat atg gat gcc tat acc ctg gcc aag gcc tac      291
Glu Glu Asp Ala Gln Asp Met Asp Ala Tyr Thr Leu Ala Lys Ala Tyr
            60              65              70
ttt gac gtt aaa gag tat gat cgg gca gca cat ttc ctg cat ggc tgc      339
Phe Asp Val Lys Glu Tyr Asp Arg Ala Ala His Phe Leu His Gly Cys
        75              80              85
aat agc aag aaa gcc tat ttt ctg tat atg tat tcc aga tat ctg gtg      387
Asn Ser Lys Lys Ala Tyr Phe Leu Tyr Met Tyr Ser Arg Tyr Leu Val
    90              95              100
agg gcc att tta aaa tgt cat tct gcc ttt agt gaa aca tcc ata ttt      435
Arg Ala Ile Leu Lys Cys His Ser Ala Phe Ser Glu Thr Ser Ile Phe
105             110             115             120
aga acc aat gga aaa gtt aaa tct ttt aaa tagcttagca gtgggccact         485
Arg Thr Asn Gly Lys Val Lys Ser Phe Lys
            125             130
gaatgaatgt actttataca tagcaataat aaaaaaaaga tatcataaat aaagttaaaa      545
aggatggtag agaagaaaat attccttagga atgactaaca ggataagtaa caacctgatt      605
atttatttac tttaggttat ataaggttct tcatgcctgt gaattaatat tattgtgtaa      665
gaattaagtt aaaaagcctg ggctgacttt aaatttata aattcattta tcatgtttat       725
agtatattta ttgttttttct ttcatggcta ttaaaaagta tgactgtaaa ggacaatgca      785
agnaaaccaa cttaatactg tattgaataa taagtacaat ttattatttt actttgaaac      845
attatgaatt tactttccta cttttttctta gttgttatct atataaattg attaaaaaaa      905
catttatgt acntnncatt tcctagtaca ggttgagtat cccttatttg a                956
```


```
<210> 100
<211> 1041
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 20..703

<221> sig_peptide
<222> 20..94
<223> Von Heijne matrix
      score 3.9
      seq ATVGLLMLGVTLP/NS

<221> polyA_signal
<222> 1000..1005

<221> polyA_site
<222> 1023..1041

<400> 100
cagggtcctg catcctacc atg tcg atg gct gtg gaa acc ttt ggc ttc ttc      52
                     Met Ser Met Ala Val Glu Thr Phe Gly Phe Phe
                     -25             -20             -15
atg gca act gtg ggg ctg ctg atg ctg ggg gtg act ctg cca aac agc      100
Met Ala Thr Val Gly Leu Leu Met Leu Gly Val Thr Leu Pro Asn Ser
        -10             -5              1
tac tgg cga gtg tcc act gtg cac ggg aac gtc atc acc acc aac acc      148
Tyr Trp Arg Val Ser Thr Val His Gly Asn Val Ile Thr Thr Asn Thr
        5               10              15
```

```
atc ttc gag aac ctc tgg ttt agc tgt gcc acc gac tcc ctg ggc gtc    196
Ile Phe Glu Asn Leu Trp Phe Ser Cys Ala Thr Asp Ser Leu Gly Val
    20              25                  30
tac aac tgc tgg gag ttc ccg tcc atg ctg gcc ctc tct ggg tat att    244
Tyr Asn Cys Trp Glu Phe Pro Ser Met Leu Ala Leu Ser Gly Tyr Ile
35              40                  45                  50
cag gcc tgc cgg gca ctc atg atc acc gcc atc ctc ctg ggc ttc ctc    292
Gln Ala Cys Arg Ala Leu Met Ile Thr Ala Ile Leu Leu Gly Phe Leu
            55                  60                  65
ggc ctc ttg cta ggc ata gcg ggc ctc cgc tgc acc aac att ggg ggc    340
Gly Leu Leu Leu Gly Ile Ala Gly Leu Arg Cys Thr Asn Ile Gly Gly
        70                  75                  80
ctg gag ctc tcc agg aaa gcc aag ctg gcg gcc acc gca ggg gcc ccc    388
Leu Glu Leu Ser Arg Lys Ala Lys Leu Ala Ala Thr Ala Gly Ala Pro
            85                  90                  95
cac att ctg gcc ggt atc tgc ggg atg gtg gcc atc tcc tgg tac gcc    436
His Ile Leu Ala Gly Ile Cys Gly Met Val Ala Ile Ser Trp Tyr Ala
    100                 105                 110
ttc aac atc acc cgg gac ttc ttc gac ccc ttg tac ccc gga acc aag    484
Phe Asn Ile Thr Arg Asp Phe Phe Asp Pro Leu Tyr Pro Gly Thr Lys
115                 120                 125                 130
tac gag ctg ggc ccc gcc ctc tac ctg ggg tgg agc gcc tca ctg atc    532
Tyr Glu Leu Gly Pro Ala Leu Tyr Leu Gly Trp Ser Ala Ser Leu Ile
            135                 140                 145
tcc atc ctg ggt ggc ctc tgc ctc tgc tcc gcc tgc tgc tgc ggc tct    580
Ser Ile Leu Gly Gly Leu Cys Leu Cys Ser Ala Cys Cys Cys Gly Ser
        150                 155                 160
gac gag gac cca gcc gcc agc gcc cgg cgg ccc tac cag gct cca gtg    628
Asp Glu Asp Pro Ala Ala Ser Ala Arg Arg Pro Tyr Gln Ala Pro Val
        165                 170                 175
tcc gtg atg ccc gtc gcc acc tcg gac caa gaa ggc gac agc agc ttt    676
Ser Val Met Pro Val Ala Thr Ser Asp Gln Glu Gly Asp Ser Ser Phe
    180                 185                 190
ggc aaa tac ggc aga aac gcc tac gtg tagcagctct ggcccgtggg         723
Gly Lys Tyr Gly Arg Asn Ala Tyr Val
195                 200
ccccgctgtc ttcccactgc cccaaggaga ggggacctgg ccggggccca ttcccctata  783
gtaacctcag gggccggcca cgccccgctc ccgtagcccc gccccggcca cggccccgtg  843
tcttgcactc tcatggcccc tccaggccaa gaactgctct tgggaagtcg catatctccc  903
ctctgaggct ggatccctca tcttctgacc ctgggttctg ggctgtgaag gggacggtgt  963
ccccgcacgt ttgtattgtg tataaataca ttcattaata aatgcatatt gtgaccgtta  1023
aaaaaaaaaa aaaaaaaa                                                 1041
```

```
<210> 101
<211> 558
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 103..294

<221> sig_peptide
<222> 103..243
<223> Von Heijne matrix
      score 5.9
      seq TWLGLLSFQNLHC/FP

<400> 101
ttcccatggt ttagaagcat aacctgtaat gtaatgcaag tcccctaact ccctggttgc    60
```

```
taacattaac ttccttaagt aataatcaat gaaagaaatt ct atg cat ggt ttt          114
                                                 Met His Gly Phe
                                                     -45
gaa ata ata tcc ttg aaa gag gaa tca cca tta gga aag gtg agt cag         162
Glu Ile Ile Ser Leu Lys Glu Glu Ser Pro Leu Gly Lys Val Ser Gln
            -40                 -35                 -30
ggt cct ttg ttt aat gtg act agt ggc tca tca tca cca gtg acc tgg         210
Gly Pro Leu Phe Asn Val Thr Ser Gly Ser Ser Ser Pro Val Thr Trp
        -25                 -20                 -15
ttg ggc cta ctc tcc ttc cag aac ctg cat tgc ttc cca gac ctc ccc         258
Leu Gly Leu Leu Ser Phe Gln Asn Leu His Cys Phe Pro Asp Leu Pro
    -10                 -5                  1                   5
act gag atg cct cta aga gcc aaa gga gtc aac act tgagcctagg              304
Thr Glu Met Pro Leu Arg Ala Lys Gly Val Asn Thr
                10                  15
gtgggctaca acaaaagatt ctaatttacc ttgcttcatc taggtccagg ccccaagtag       364
cttgctgaag gaacttaaaa agtagctgtt atttattgta ttgtataagc taaaaacatt       424
tatttttgtt gaatcgaaac aattccatgt agcaatcttt tttctgttca cggtgttcgt       484
gatagaacct taaattccgc aagcatcagt tttttgaaaa aatgggaatt gaccggatag       544
taacaggcaa agtt                                                         558
```

```
<210> 102
<211> 730
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 81..518

<221> sig_peptide
<222> 81..173
<223> Von Heijne matrix
      score 3.9
      seq ILFHGVFYAGGFA/IV

<400> 102
ctcgtcatgc tctttgtagc gtggtgcttc tgttgctcac aggacaactt gcctttgatg      60
attttcaaga gagttgtgct atg atg tgg caa aag tat gca gga agc agg cgg      113
                      Met Met Trp Gln Lys Tyr Ala Gly Ser Arg Arg
                          -30                 -25
tca atg cct ctg gga gca agg atc ctt ttc cac ggt gtg ttc tat gcc        161
Ser Met Pro Leu Gly Ala Arg Ile Leu Phe His Gly Val Phe Tyr Ala
-20                 -15                 -10                 -5
ggg ggc ttt gcc att gtg tat tac ctc att caa aag ttt cat tcc agg        209
Gly Gly Phe Ala Ile Val Tyr Tyr Leu Ile Gln Lys Phe His Ser Arg
                1                   5                   10
gct tta tat tac aag ttg gca gtg gag cag ctg cag agc cat ccc gag        257
Ala Leu Tyr Tyr Lys Leu Ala Val Glu Gln Leu Gln Ser His Pro Glu
            15                  20                  25
gca cag gaa gct ctg ggc cct cct ctc aac atc cat tat ctc aag ctc        305
Ala Gln Glu Ala Leu Gly Pro Pro Leu Asn Ile His Tyr Leu Lys Leu
        30                  35                  40
atc gac agg gaa aac ttc gtg gac att gtt gat gcc aag ttg aag att        353
Ile Asp Arg Glu Asn Phe Val Asp Ile Val Asp Ala Lys Leu Lys Ile
45                  50                  55                  60
cct gtc tct gga tcc aaa tca gag ggc ctt ctc tac gtc cac tca tcc        401
Pro Val Ser Gly Ser Lys Ser Glu Gly Leu Leu Tyr Val His Ser Ser
                65                  70                  75
aga ggt ggc ccc ttt cag agg tgg cac ctt gac gag gtc ttt tta gag       449
```

```
Arg Gly Gly Pro Phe Gln Arg Trp His Leu Asp Glu Val Phe Leu Glu
        80                  85                  90
ctc aag gat ggt cag cag att cct gtg ttc aag ctc agt ggg gaa aac    497
Leu Lys Asp Gly Gln Gln Ile Pro Val Phe Lys Leu Ser Gly Glu Asn
        95                 100                 105
ggt gat gaa gtg aaa aag gag tagagacgac ccagaagacc cagcttgctt       548
Gly Asp Glu Val Lys Lys Glu
        110                 115
ctagtccatc cttccctcat ctctaccata tggccactgg ggtggtggcc catctcagtg  608
acagacactc ctgcaaccca gttttccagc caccagtggg atgatggtat gtgccagcac  668
atggtaattt tggtgtaatt ctaacttggg cacaacgaat gctatttgtc atttttaaac  728
tg                                                                 730
```

```
<210> 103
<211> 1098
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 66..326

<221> polyA_signal
<222> 1066..1071

<221> polyA_site
<222> 1087..1098

<400> 103
ctcccttttga atgagagaaa ctaacccgct tccgaagccc ctgaaagaca ctgctccttc   60
ctctc atg gag ttg gct ccg aca gcc cgt ctg cca cca ggc cat ggt tcc  110
      Met Glu Leu Ala Pro Thr Ala Arg Leu Pro Pro Gly His Gly Ser
          1           5                   10                  15
ttg ccc cat ggt gtc ctg gga ccc aga gca aca gga tct gtc acc cac    158
Leu Pro His Gly Val Leu Gly Pro Arg Ala Thr Gly Ser Val Thr His
                20                  25                  30
ctc tct ctt ctc ccc cag atc aag caa cgt gcc tca gag gct ttg ccc    206
Leu Ser Leu Leu Pro Gln Ile Lys Gln Arg Ala Ser Glu Ala Leu Pro
            35                  40                  45
gaa ttg ctt cgt cct gtc acc ccc atc acc aat ttt gag ggc agc cag    254
Glu Leu Leu Arg Pro Val Thr Pro Ile Thr Asn Phe Glu Gly Ser Gln
        50                  55                  60
tct cag gac cac agt gga atc ttt ggc ctg gta aca aac ctg gaa gag    302
Ser Gln Asp His Ser Gly Ile Phe Gly Leu Val Thr Asn Leu Glu Glu
    65                  70                  75
ctg gag gtg gac gat tgg gag ttc tgagcctctg caaactgtgc gcattctcca   356
Leu Glu Val Asp Asp Trp Glu Phe
80                  85
gccagggatg cagaggccac ccagaggccc ttcctgaggg ccggccacat tcccgccctc   416
ctgggcagat tgggtagaaa ggacattctt ccaggaaagt tgactgctgg ctgattggga   476
aagaaaatcc tggagagata cttcactgct ccaaggcttt tgagacacaa gggaatctca   536
acaaccaggg atcaggaggg tccaaagccg acattcccag tcctgtgagc tcaggtgacc   596
tcctccgcag aagagagatg ctgctctggc cctgggagct gaattccaag cccagggttt   656
ggctccttaa acccgaggac cgccacctct tcccagtgct gcgaccagc ctcattctac    716
ttaactttgc tctcagatgc ctcagatgct ataggtcagt gaaagggcga gtagtaagct   776
gcctgcctcc cttccctcag acctctccct cataattcca gagaagggca tttctgtctt   836
tttaagcaca gactaaggct ggaacagtcc atcctatcc ctcttctggc ttgggccctg    896
acacctaagt cttttcccacg gtttatgtgt gtgcctcatt cctttcccac caagaatcca   956
tcttagcgcc tcctgccagc tgccctggtg ctttctcaa gggcatcag tgtcttgcct     1016
agcttgaggg cttaagtcct tatgctgtgt tagtttcgtt gtcagaacaa attaaaattt   1076
```

tcagagacgc aaaaaaaaaa aa   1098

<210> 104
<211> 346
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 170..289

<221> sig_peptide
<222> 170..250
<223> Von Heijne matrix
      score 3.6
      seq LTLLLITPSPSPL/LF

<400> 104
ccatttgagc cccaccacgg aggttatgtg gtcccaaaag gaatgatggc caagcaatta   60
attttttcctc ctagttctta gcttgcttct gcattgattg gctttacaca actggcattt   120
agtctgcatt acacaaatag acactaattt atttggaaca agcagcaaa atg aga act   178
                                                        Met Arg Thr
                                                             -25
tta ttt ggt gca gtc agg gct cca ttt agt tcc ctc act ctg ctt cta   226
Leu Phe Gly Ala Val Arg Ala Pro Phe Ser Ser Leu Thr Leu Leu Leu
         -20   -15   -10
atc acc cct tct ccc agc cct ctt cta ttt gat aga ggt ctg tcc ctc   274
Ile Thr Pro Ser Pro Ser Pro Leu Leu Phe Asp Arg Gly Leu Ser Leu
              -5   1   5
aga tca gca atg tct tagcccctct cctctcttcc attccttcct gttggtactc   329
Arg Ser Ala Met Ser
         10
atttcttcta actttta   346

<210> 105
<211> 685
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 36..497

<221> polyA_signal
<222> 650..655

<221> polyA_site
<222> 663..685

<400> 105
aagttctgcg ctggtcggcg gagtagcaag tggcc atg ggg agc ctc agc ggt   53
                                        Met Gly Ser Leu Ser Gly
                                          1   5
ctg cgc ctg gca gca gga agc tgt ttt agg tta tgt gaa aga gat gtt   101
Leu Arg Leu Ala Ala Gly Ser Cys Phe Arg Leu Cys Glu Arg Asp Val
         10   15   20
tcc tca tct cta agg ctt acc aga agc tct gat ttg aag aga ata aat   149
Ser Ser Ser Leu Arg Leu Thr Arg Ser Ser Asp Leu Lys Arg Ile Asn

```
          25                    30                    35
gga ttt tgc aca aaa cca cag gaa agt ccc gga gct cca tcc cgc act      197
Gly Phe Cys Thr Lys Pro Gln Glu Ser Pro Gly Ala Pro Ser Arg Thr
    40                    45                    50
tac aac aga gtg cct tta cac aaa cct acg gat tgg cag aaa aag atc      245
Tyr Asn Arg Val Pro Leu His Lys Pro Thr Asp Trp Gln Lys Lys Ile
55                    60                    65                    70
ctc ata tgg tca ggt cgc ttc aaa aag gaa gat gaa atc cca gag act      293
Leu Ile Trp Ser Gly Arg Phe Lys Lys Glu Asp Glu Ile Pro Glu Thr
                  75                    80                    85
gtc tcg ttg gag atg ctt gat gct gca aag aac aag atg cga gtg aag      341
Val Ser Leu Glu Met Leu Asp Ala Ala Lys Asn Lys Met Arg Val Lys
                  90                    95                    100
agc agc tat cta atg att gcc ctg acg gtg gta gga tgc atc ttc atg      389
Ser Ser Tyr Leu Met Ile Ala Leu Thr Val Val Gly Cys Ile Phe Met
          105                   110                   115
gtt att gag ggc aag aag gct gcc caa aga cac gag act tta aca agc      437
Val Ile Glu Gly Lys Lys Ala Ala Gln Arg His Glu Thr Leu Thr Ser
          120                   125                   130
ttg aac tta gaa aag aaa gct cgt ctg aaa gag gaa gca gct atg aag      485
Leu Asn Leu Glu Lys Lys Ala Arg Leu Lys Glu Glu Ala Ala Met Lys
135                   140                   145                   150
gcc aaa aca gag tagcagaggt atccgtgttg gctggatttt gaaaatccag          537
Ala Lys Thr Glu
gaattatgtt ataacgtgcc tgtattaaaa aggatgtggt atgaggatcc atttcataaa    597
gtatgatttg cccaaacctg taccatttcc gtatttctgc cgtagaagta gaaataaatt    657
ttcttaaaaa aaaaaaaaaa aaaaaaaa                                        685
```

```
<210> 106
<211> 554
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 18..320

<221> polyA_signal
<222> 539..544

<221> polyA_site
<222> 542..554

<400> 106
aaccgtcgtg gggaagg atg gtg tgc gaa aaa tgt gaa aag aaa ctt ggt       50
                    Met Val Cys Glu Lys Cys Glu Lys Lys Leu Gly
                    1                   5                   10
act gtt atc act cca gat aca tgg aaa gat ggt gct agg aat acc aca      98
Thr Val Ile Thr Pro Asp Thr Trp Lys Asp Gly Ala Arg Asn Thr Thr
                  15                    20                    25
gaa agt ggt gga aga aag ctg aat aaa aat aaa gct ttg act tca aaa      146
Glu Ser Gly Gly Arg Lys Leu Asn Lys Asn Lys Ala Leu Thr Ser Lys
          30                    35                    40
aaa gca aga ttt gat cca tat gga aag aat aag ttc tcc act tgt aga      194
Lys Ala Arg Phe Asp Pro Tyr Gly Lys Asn Lys Phe Ser Thr Cys Arg
          45                    50                    55
att tgt aaa agt tct gtg cac caa cca ggt tct cat tac tgc cag ggc      242
Ile Cys Lys Ser Ser Val His Gln Pro Gly Ser His Tyr Cys Gln Gly
60                    65                    70                    75
tgt gcc tac aaa aaa ggc atc tgt gcg atg tgt ggn aaa aaa gtt ttg      290
```

```
Cys Ala Tyr Lys Lys Gly Ile Cys Ala Met Cys Gly Lys Lys Val Leu
              80                  85                  90
gat acc aaa aac tac aag caa aca tct gtc tagatgtatt gatggaattt       340
Asp Thr Lys Asn Tyr Lys Gln Thr Ser Val
              95                  100
ctggctttct aaatgatttt actttctgcc ttgaattttc aaggcataga tgtcaactta   400
cagaataaca tgttttaaga taattaagtt taaaccagag aatttgattg ttactcattt   460
tgctctcatg ttctaaacag caacagtgta actagtcttt tgttgtaaat ggttattttc   520
cttataagaa ttttaagaac taaaaaaaaa aaaa                                554
```

<210> 107
<211> 1678
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 71..1438

<221> sig_peptide
<222> 71..136
<223> Von Heijne matrix
      score 3.5
      seq AAPVAAGLGPVIS/RP

<221> polyA_signal
<222> 1644..1649

<221> polyA_site
<222> 1665..1678

<400> 107
```
ccgacttcca gaggagcgct gtgcacgtgg agaagagcgg ggactcggcg accctgccct   60
cccgaccctc atg ttc gaa gag cct gag tgg gcc gag gcg gcc cca gta       109
            Met Phe Glu Glu Pro Glu Trp Ala Glu Ala Ala Pro Val
                -20                 -15                 -10
gcc gcg ggc ctt ggg ccc gta atc tca cga cct ccg cct gcg gcc tcc     157
Ala Ala Gly Leu Gly Pro Val Ile Ser Arg Pro Pro Pro Ala Ala Ser
        -5                   1                   5
tcg caa aac aag ggc tcc aag cgc cgc cag ctc ttg gcc aca tta cgg     205
Ser Gln Asn Lys Gly Ser Lys Arg Arg Gln Leu Leu Ala Thr Leu Arg
        10                  15                  20
gcc cta gag gca gca tct ctt tcc cag cat ccc ccc agc cta tgt ata     253
Ala Leu Glu Ala Ala Ser Leu Ser Gln His Pro Pro Ser Leu Cys Ile
    25                  30                  35
agt gac tct gag gag gag gag gag gaa agg aag aag aaa tgc ccc aaa     301
Ser Asp Ser Glu Glu Glu Glu Glu Glu Arg Lys Lys Lys Cys Pro Lys
40                  45                  50                  55
aag gca tca ttt gcc agt gcc tct gct gaa gta ggg aag aaa ggg aag     349
Lys Ala Ser Phe Ala Ser Ala Ser Ala Glu Val Gly Lys Lys Gly Lys
            60                  65                  70
aag aaa tgt caa aaa cag ggc cca cct tgc agt gac tct gag gaa gaa     397
Lys Lys Cys Gln Lys Gln Gly Pro Pro Cys Ser Asp Ser Glu Glu Glu
            75                  80                  85
gta gaa agg aag aag aaa tgc cac aaa cag gct ctt gtt ggc agt gac     445
Val Glu Arg Lys Lys Lys Cys His Lys Gln Ala Leu Val Gly Ser Asp
            90                  95                  100
tct gct gaa gat gag aaa aga aag agg aaa tgc cag aaa cat gcc cct     493
Ser Ala Glu Asp Glu Lys Arg Lys Arg Lys Cys Gln Lys His Ala Pro
    105                 110                 115
```

```
ata aat tca gcc cag cac ctg gac aat gtt gac caa aca ggt ccc aaa          541
Ile Asn Ser Ala Gln His Leu Asp Asn Val Asp Gln Thr Gly Pro Lys
120                 125                 130                 135
gcc tgg aag ggt agt act aca aat gat cca cca aag caa agc cct ggg          589
Ala Trp Lys Gly Ser Thr Thr Asn Asp Pro Pro Lys Gln Ser Pro Gly
                140                 145                 150
tcc act tcc cct aaa ccc cct cat aca tta agc cgc aag cag tgg cgg          637
Ser Thr Ser Pro Lys Pro Pro His Thr Leu Ser Arg Lys Gln Trp Arg
                155                 160                 165
aac cgg caa aag aat aag aga aga tgt aag aac aag ttt cag cca cct          685
Asn Arg Gln Lys Asn Lys Arg Arg Cys Lys Asn Lys Phe Gln Pro Pro
                170                 175                 180
cag gtg cca gac cag gcc cca gct gag gcc ccc aca gag aag aca gag          733
Gln Val Pro Asp Gln Ala Pro Ala Glu Ala Pro Thr Glu Lys Thr Glu
        185                 190                 195
gtg tct cct gtt ccc agg aca gac agc cat ggg gct cgg gca ggg gct          781
Val Ser Pro Val Pro Arg Thr Asp Ser His Gly Ala Arg Ala Gly Ala
200                 205                 210                 215
ttg cga gcc cgc atg gca cag cgg ctg gat ggg gcc cga ttt cgc tac          829
Leu Arg Ala Arg Met Ala Gln Arg Leu Asp Gly Ala Arg Phe Arg Tyr
                220                 225                 230
ctc aat gaa cag ttg tac tca ggg ccc agc agt gct gca cag cgt ctc          877
Leu Asn Glu Gln Leu Tyr Ser Gly Pro Ser Ser Ala Ala Gln Arg Leu
                235                 240                 245
ttc cag gaa gac cct gag gct ttt ctt ctc tac cac cgc ggc ttc cag          925
Phe Gln Glu Asp Pro Glu Ala Phe Leu Leu Tyr His Arg Gly Phe Gln
        250                 255                 260
agc caa gtg aag aag tgg cca ctg cag cca gtg gac cgc atc gcc agg          973
Ser Gln Val Lys Lys Trp Pro Leu Gln Pro Val Asp Arg Ile Ala Arg
        265                 270                 275
gat ctt cgc cag cgg cct gca tcc cta gtg gtg gct gac ttc ggc tgt          1021
Asp Leu Arg Gln Arg Pro Ala Ser Leu Val Val Ala Asp Phe Gly Cys
280                 285                 290                 295
ggg gat tgc cgc ttg gct tca agt atc cgg aac cct gtg cat tgc ttt          1069
Gly Asp Cys Arg Leu Ala Ser Ser Ile Arg Asn Pro Val His Cys Phe
                300                 305                 310
gac ttg gct tct ctg gac cct agg gtc act gtg tgt gac atg gcc cag          1117
Asp Leu Ala Ser Leu Asp Pro Arg Val Thr Val Cys Asp Met Ala Gln
                315                 320                 325
gtt cct ttg gag gat gag tct gtg gat gtg gct gtg ttt tgc ctt tca          1165
Val Pro Leu Glu Asp Glu Ser Val Asp Val Ala Val Phe Cys Leu Ser
        330                 335                 340
ctg atg gga acc aac atc agg gac ttc cta gag gag gca aat aga gta          1213
Leu Met Gly Thr Asn Ile Arg Asp Phe Leu Glu Glu Ala Asn Arg Val
        345                 350                 355
ctg aag cca ggg ggt ctc ctg aaa gtg gct gag gtc agc agc cgc ttt          1261
Leu Lys Pro Gly Gly Leu Leu Lys Val Ala Glu Val Ser Ser Arg Phe
360                 365                 370                 375
gag gat gtt cga acc ttt ctg cgg gct gtg acc aag cta ggc ttc aag          1309
Glu Asp Val Arg Thr Phe Leu Arg Ala Val Thr Lys Leu Gly Phe Lys
                380                 385                 390
att gtc tcc aag gac ctg acc aac agc cat ttc ttc ttg ttt gat ttc          1357
Ile Val Ser Lys Asp Leu Thr Asn Ser His Phe Phe Leu Phe Asp Phe
                395                 400                 405
caa aag act ggg ccc cct ctg gta ggg ccc aag gct cag ctt tca ggc          1405
Gln Lys Thr Gly Pro Pro Leu Val Gly Pro Lys Ala Gln Leu Ser Gly
                410                 415                 420
ctg cag ctt cag cca tgt ctc tac aag cgc agg tgacctctgg atcttccttg        1458
Leu Gln Leu Gln Pro Cys Leu Tyr Lys Arg Arg
                425                 430
agaggggagg cagatctcaa actccaggct cagaactgtg aagactgttt ccggcctggc       1518
tgtgagccaa gacctggttc ctggtggacc ctgaggacaa agtgtgataa aacctctggc       1578
```

tcagacttgc tctactgaag gcttcttggt tataagatgc ataaagtcac tggggctagc   1638
taaacaataa agagtttatt gtgaggaaaa aaaaaaaaaa   1678

<210> 108
<211> 494
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 25..318

<221> sig_peptide
<222> 25..75
<223> Von Heijne matrix
      score 7.4
      seq FFLLLQFFLRIDG/VL

<221> polyA_signal
<222> 452..457

<221> polyA_site
<222> 482..494

<400> 108
aggctgagtg tgaagattag agta atg cct tct agc ttt ttc ctg ctg ttg   51
                           Met Pro Ser Ser Phe Phe Leu Leu Leu
                               -15   -10
cag ttt ttc ctg aga att gat ggg gtg ctt atc aga atg aat gac acg   99
Gln Phe Phe Leu Arg Ile Asp Gly Val Leu Ile Arg Met Asn Asp Thr
        -5   1   5
aga ctt tac cat gag gct gac aag acc tac atg tta cga gaa tat acg   147
Arg Leu Tyr His Glu Ala Asp Lys Thr Tyr Met Leu Arg Glu Tyr Thr
    10   15   20
tca cga gaa agc aaa att tct agt ttg atg cat gtt cca cct tcc ctc   195
Ser Arg Glu Ser Lys Ile Ser Ser Leu Met His Val Pro Pro Ser Leu
25   30   35   40
ttc acg gaa cct aat gaa ata tcc cag tat tta cca ata aag gaa gca   243
Phe Thr Glu Pro Asn Glu Ile Ser Gln Tyr Leu Pro Ile Lys Glu Ala
                45   50   55
gtt tgt gag aag cta ata ttt cca gaa aga att gat cct aac cca gca   291
Val Cys Glu Lys Leu Ile Phe Pro Glu Arg Ile Asp Pro Asn Pro Ala
            60   65   70
gac tca caa aaa agt aca caa gtg gaa taaaatgtga tacaacatat   338
Asp Ser Gln Lys Ser Thr Gln Val Glu
        75   80
actcactatg gaatctgact ggacaccttg gctatttgta aggggttatt tttattatga   398
gaattaattg ccttgtttat gtacagattt tctgtagcct taaaggaaaa aaaaataaag   458
atcgttacag gcaggtttca ctcaaaaaaa aaaaac   494

<210> 109
<211> 714
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 84..332

```
<221> sig_peptide
<222> 84..170
<223> Von Heijne matrix
      score 5.2
      seq PCYYLGLFQRALA/SV

<221> polyA_site
<222> 702..714

<400> 109
cctatctctt ctgctggctg ggctcaatgc cgcgggtgag cgttcggccg aggctgctcc    60
tacccttgag tgatgtgcct tga atg acg ctg ctt tca ttc gct gct ttc acg   113
                          Met Thr Leu Leu Ser Phe Ala Ala Phe Thr
                              -25                    -20
gct gct ttc tcc gtc ctc ccc tgt tac tac ctt ggg ctg ttt cag cgg    161
Ala Ala Phe Ser Val Leu Pro Cys Tyr Tyr Leu Gly Leu Phe Gln Arg
            -15             -10                     -5
gcg ctc gcg tcg gtc ttc gac cca ctt tgc gtt tgt tca cgt gtg ctc    209
Ala Leu Ala Ser Val Phe Asp Pro Leu Cys Val Cys Ser Arg Val Leu
        1           5                   10
ccg aca cct gta tgt acc ttg gtc gca aca caa gcc gaa aaa ata tta    257
Pro Thr Pro Val Cys Thr Leu Val Ala Thr Gln Ala Glu Lys Ile Leu
    15              20              25
gag aat ggg ccc tgt cca acc aag gag gcg gcc cag ctt gtc ggg aag    305
Glu Asn Gly Pro Cys Pro Thr Lys Glu Ala Ala Gln Leu Val Gly Lys
30              35              40                  45
ggc agc gtt tcc gcc aga aat gct tcg tgaaaggcac ttgagggacc          352
Gly Ser Val Ser Ala Arg Asn Ala Ser
            50
ttagcagcat cctcaacagg ccttgtaggg aatgccagaa gaagcagtcc ttggccgggc   412
ggggtggctc atgcctgtgg tcccagcact ttgggaggcc ggggcgggcg gatcacctga   472
ggtcgggagg tccagaccag cctgaccgac atggagaaac cccgtctnta ctagaaatac   532
aaaactagcc gggtgtggtg gcgcatgcct gtagtcccag ctactcggga gggtgaggca   592
ggagacgttc ttgaacccgg gaggcggagt ttgtggtgag ccgagatcgc gccattgcac   652
tccagcctgg gcatgccaag agcgaaactc cgtctcaaaa aaaaaaaga aaaaaaaaa     712
aa                                                                  714


<210> 110
<211> 805
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 32..718

<221> sig_peptide
<222> 32..100
<223> Von Heijne matrix
      score 7.4
      seq VLLLAALPPVLLP/GA

<221> polyA_signal
<222> 770..775

<221> polyA_site
<222> 793..805

<400> 110
```

```
cctctttcag cccgggatcg ccccagcagg g atg ggc gac aag atc tgg ctg          52
                                    Met Gly Asp Lys Ile Trp Leu
                                                           -20
ccc ttc ccc gtg ctc ctt ctg gcc gct ctg cct ccg gtg ctg ctg cct        100
Pro Phe Pro Val Leu Leu Leu Ala Ala Leu Pro Pro Val Leu Leu Pro
    -15                     -10                     -5
ggg gcg gcc ggc ttc aca cct tcc ctc gat agc gac ttc acc ttt acc        148
Gly Ala Ala Gly Phe Thr Pro Ser Leu Asp Ser Asp Phe Thr Phe Thr
1               5                   10                  15
ctt ccc gcc ggc cag aag gag tgc ttc tac cag ccc atg ccc ctg aag        196
Leu Pro Ala Gly Gln Lys Glu Cys Phe Tyr Gln Pro Met Pro Leu Lys
                20                  25                  30
gcc tcg ctg gag atc gag tac caa gtt tta gat gga gca gga tta gat        244
Ala Ser Leu Glu Ile Glu Tyr Gln Val Leu Asp Gly Ala Gly Leu Asp
            35                  40                  45
att gat ttc cat ctt gcc tct cca gaa ggc aaa acc tta gtt ttt gaa        292
Ile Asp Phe His Leu Ala Ser Pro Glu Gly Lys Thr Leu Val Phe Glu
        50                  55                  60
caa aga aaa tca gat gga gtt cac act gta gag act gaa gtt ggt gat        340
Gln Arg Lys Ser Asp Gly Val His Thr Val Glu Thr Glu Val Gly Asp
65                  70                  75                  80
tac atg ttc tgc ttt gac aat aca ttc agc acc att tct gag aag gtg        388
Tyr Met Phe Cys Phe Asp Asn Thr Phe Ser Thr Ile Ser Glu Lys Val
                85                  90                  95
att ttc ttt gaa tta atc ctg gat aat atg gga gaa cag gca caa gaa        436
Ile Phe Phe Glu Leu Ile Leu Asp Asn Met Gly Glu Gln Ala Gln Glu
                100                 105                 110
caa gaa gat tgg aag aaa tat att act ggc aca gat ata ttg gat atg        484
Gln Glu Asp Trp Lys Lys Tyr Ile Thr Gly Thr Asp Ile Leu Asp Met
            115                 120                 125
aaa ctg gaa gac atc ctg gaa tcc atc agc agc atc aag tcc aga cta        532
Lys Leu Glu Asp Ile Leu Glu Ser Ile Ser Ser Ile Lys Ser Arg Leu
        130                 135                 140
agc aaa agt ggg cac ata caa att ctg ctt aga gca ttt gaa gct cgt        580
Ser Lys Ser Gly His Ile Gln Ile Leu Leu Arg Ala Phe Glu Ala Arg
145                 150                 155                 160
gat cga aac ata caa gaa agc aac ttt gat aga gtc aat ttc tgg tct        628
Asp Arg Asn Ile Gln Glu Ser Asn Phe Asp Arg Val Asn Phe Trp Ser
                165                 170                 175
atg gtt aat tta gtg gtc atg gtg gtg gtg tca gcc att caa gtt tat        676
Met Val Asn Leu Val Val Met Val Val Val Ser Ala Ile Gln Val Tyr
                180                 185                 190
atg ctg aag agt ctg ttt gaa gat aag agg aaa agt aga act           718
Met Leu Lys Ser Leu Phe Glu Asp Lys Arg Lys Ser Arg Thr
        195                 200                 205
taaaactcca aactagagta cgtaacattg aaaaatgagg cataaaaatg caataaactg        778
ttacagtcaa gaccaaaaaa aaaaaaa                                            805
```

<210> 111
<211> 787
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 26..481

<221> sig_peptide
<222> 26..88
<223> Von Heijne matrix

score 4.4
seq AVASSFFCASLFS/AV

<221> polyA_signal
<222> 755..760

<221> polyA_site
<222> 775..787

<400> 111
```
gacagcctgg ataaaggctc acttg atg gct cag ttg gga gca gtt gtg gct        52
                            Met Ala Gln Leu Gly Ala Val Val Ala
                                -20                 -15
gtg gct tcc agt ttc ttt tgt gca tct ctc ttc tca gct gtg cac aag       100
Val Ala Ser Ser Phe Phe Cys Ala Ser Leu Phe Ser Ala Val His Lys
        -10             -5                  1
ata gaa gag gga cat att ggg gta tat tac aga ggc ggt gcc ctg ctg       148
Ile Glu Glu Gly His Ile Gly Val Tyr Tyr Arg Gly Gly Ala Leu Leu
5               10              15                  20
act tcg acc agc ggc cct ggt ttc cat ctc atg ctc cct ttc atc aca       196
Thr Ser Thr Ser Gly Pro Gly Phe His Leu Met Leu Pro Phe Ile Thr
            25              30                  35
tca tat aag tct gtg cag acc aca ctc cag aca gat gag gtg aag aat       244
Ser Tyr Lys Ser Val Gln Thr Thr Leu Gln Thr Asp Glu Val Lys Asn
        40              45                  50
gta cct tgt ggg act agt ggt ggt gtg atg atc tac ttt gac aga att       292
Val Pro Cys Gly Thr Ser Gly Gly Val Met Ile Tyr Phe Asp Arg Ile
        55              60                  65
gaa gtg gtg aac ttc ctg gtc ccg aac gca gtg cat gat ata gtg aag       340
Glu Val Val Asn Phe Leu Val Pro Asn Ala Val His Asp Ile Val Lys
70              75                  80
aac tat act gct gac tat gac aag gcc ctc atc ttc aac aag atc cac       388
Asn Tyr Thr Ala Asp Tyr Asp Lys Ala Leu Ile Phe Asn Lys Ile His
85              90                  95                  100
cac gaa ctg aac cag ttc tgc agt gtg cac acg ctt caa gag gtc tac       436
His Glu Leu Asn Gln Phe Cys Ser Val His Thr Leu Gln Glu Val Tyr
            105             110                 115
att gag ctg ttt gga ctg gaa aat gat ttt tcc cag gaa tct tca           481
Ile Glu Leu Phe Gly Leu Glu Asn Asp Phe Ser Gln Glu Ser Ser
            120             125                 130
taaaagggac cctgagcaag aacatttttc atagcagaca ggaggactca tccacatcgc      541
cagcaatcat aattaagcaa accgcctttt gcaccattta agatttagga aatcatccaa      601
attactttta atgtttctgc agtagaaaat gaatctaaat tcattttata gggtttgtag      661
tcttttatct gttttggatt cactgtgctt ttaagaaaaa gttggtaaat ttgccgttga      721
tttttctttt taacctcaaa ctaatagaat tttataaaat attaattttc tccaaaaaaa      781
aaaaaa                                                                  787
```

<210> 112
<211> 569
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 26..562

<221> sig_peptide
<222> 26..187
<223> Von Heijne matrix
      score 4.1

seq AVVAAAARTGSEA/RV

<400> 112

```
agaaacaggt ctgggctaca aaagt atg gcc gct tct gag gcg gcg gtg gtg          52
                            Met Ala Ala Ser Glu Ala Ala Val Val
                                                        -50
tct tcg ccg tct ttg aaa aca gac aca tcc cct gtc ctt gaa act gca         100
Ser Ser Pro Ser Leu Lys Thr Asp Thr Ser Pro Val Leu Glu Thr Ala
-45                 -40                 -35                 -30
gga acg gtc gca gca atg gct gcg acc ccg tca gca agg gct gca gcc        148
Gly Thr Val Ala Ala Met Ala Ala Thr Pro Ser Ala Arg Ala Ala Ala
            -25                 -20                 -15
gcg gtg gtt gcg gcc gcg gcc agg acc gga tcc gaa gcc agg gtc tcc        196
Ala Val Val Ala Ala Ala Ala Arg Thr Gly Ser Glu Ala Arg Val Ser
            -10                 -5                  1
aag gcc gct ttg gct acc aag ctg ctg tcc ttg agc ggc gtg ttc gcc        244
Lys Ala Ala Leu Ala Thr Lys Leu Leu Ser Leu Ser Gly Val Phe Ala
    5                   10                  15
gtg cac aag ccc aaa ggg ccc act tca gcc gag ctg ctg aat cgg ttg        292
Val His Lys Pro Lys Gly Pro Thr Ser Ala Glu Leu Leu Asn Arg Leu
20                  25                  30                  35
aag gag aag ctg ctg gca gaa gct gga atg cct tct cca gaa tgg acc        340
Lys Glu Lys Leu Leu Ala Glu Ala Gly Met Pro Ser Pro Glu Trp Thr
            40                  45                  50
aag agg aaa aag cag act ttg aaa att ggg cat gga ggg act cta gac        388
Lys Arg Lys Lys Gln Thr Leu Lys Ile Gly His Gly Gly Thr Leu Asp
        55                  60                  65
agc gca gcc cga gga gtt ctg gtt gtt gga att gga agc gga aca aaa        436
Ser Ala Ala Arg Gly Val Leu Val Val Gly Ile Gly Ser Gly Thr Lys
        70                  75                  80
atg ttg acc agt atg ttg tca ggg tcc aag agg tat act gcc att gga        484
Met Leu Thr Ser Met Leu Ser Gly Ser Lys Arg Tyr Thr Ala Ile Gly
    85                  90                  95
gaa ctg ggg aaa gct act gat aca cta gat tct acg ggg aag gta aca        532
Glu Leu Gly Lys Ala Thr Asp Thr Leu Asp Ser Thr Gly Lys Val Thr
100                 105                 110                 115
gaa gaa aaa cct tac ggt atg aac ctc atc taagtag                        569
Glu Glu Lys Pro Tyr Gly Met Asn Leu Ile
                120                 125
```

<210> 113
<211> 893
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 4..810

<221> sig_peptide
<222> 4..279
<223> Von Heijne matrix
      score 6.8
      seq AVMLYTWRSCSRA/IP

<221> polyA_signal
<222> 858..863

<221> polyA_site
<222> 881..893

```
<400> 113
gcc atg atc acg cac gtc acc ctg gaa gat gcc ctg tcc aac gtg gac     48
    Met Ile Thr His Val Thr Leu Glu Asp Ala Leu Ser Asn Val Asp
        -90               -85                   -80
ctg ctt gaa gag ctt ccc ctc ccc gac cag cag cca tgc atc gag cct     96
Leu Leu Glu Glu Leu Pro Leu Pro Asp Gln Gln Pro Cys Ile Glu Pro
    -75               -70                   -65
cca cct tcc tcc atc atg tac cag gct aac ttt gac aca aac ttt gag    144
Pro Pro Ser Ser Ile Met Tyr Gln Ala Asn Phe Asp Thr Asn Phe Glu
    -60               -55                   -50
gac agg aat gca ttt gtc acg ggc att gca agg tac att gag cag gct    192
Asp Arg Asn Ala Phe Val Thr Gly Ile Ala Arg Tyr Ile Glu Gln Ala
-45               -40                   -35                   -30
aca gtc cac tcc agc atg aat gag atg ctg gag gaa gga cat gag tat    240
Thr Val His Ser Ser Met Asn Glu Met Leu Glu Glu Gly His Glu Tyr
            -25               -20                   -15
gcg gtc atg ctg tac acc tgg cgc agc tgt tcc cgg gcc att ccc cag    288
Ala Val Met Leu Tyr Thr Trp Arg Ser Cys Ser Arg Ala Ile Pro Gln
        -10               -5                    1
gtg aaa tgc aac gag cag ccc aac cga gta gag atc tat gag aag aca    336
Val Lys Cys Asn Glu Gln Pro Asn Arg Val Glu Ile Tyr Glu Lys Thr
    5                 10                    15
gta gag gtg ctg gag ccg gag gtc acc aag ctc atg aag ttc atg tat    384
Val Glu Val Leu Glu Pro Glu Val Thr Lys Leu Met Lys Phe Met Tyr
20                25                    30                    35
ttt cag cgc aag gcc atc gag cgg ttc tgc agc gag gtg aag cgg ctg    432
Phe Gln Arg Lys Ala Ile Glu Arg Phe Cys Ser Glu Val Lys Arg Leu
            40                45                    50
tgc cat gcc gag cgc agg aag gac ttt gtc tct gag gcc tac ctc ctg    480
Cys His Ala Glu Arg Arg Lys Asp Phe Val Ser Glu Ala Tyr Leu Leu
            55                60                    65
acc ctt ggc aag ttc atc aac atg ttt gct gtc ctg gat gag cta aag    528
Thr Leu Gly Lys Phe Ile Asn Met Phe Ala Val Leu Asp Glu Leu Lys
        70                75                    80
aac atg aag tgc agc gtc aag aat gac cac tcc gcc tac aag agg gca    576
Asn Met Lys Cys Ser Val Lys Asn Asp His Ser Ala Tyr Lys Arg Ala
        85                90                    95
gca cag ttc ctg cgg aag atg gca gat ccc cag tct atc cag gag tcg    624
Ala Gln Phe Leu Arg Lys Met Ala Asp Pro Gln Ser Ile Gln Glu Ser
100               105                   110                   115
cag aac ctt tcc atg ttc ctg gcc aac cac aac agg atc acc cag tgt    672
Gln Asn Leu Ser Met Phe Leu Ala Asn His Asn Arg Ile Thr Gln Cys
            120               125                   130
ctc cac cag caa ctt gaa gtg atc cca ggc tat gag gag ctg ctg gct    720
Leu His Gln Gln Leu Glu Val Ile Pro Gly Tyr Glu Glu Leu Leu Ala
            135               140                   145
gac att gtc aac atc tgt gtg gat tac tac gag aac aag atg tac ctg    768
Asp Ile Val Asn Ile Cys Val Asp Tyr Tyr Glu Asn Lys Met Tyr Leu
            150               155                   160
act ccc agt gag aaa cat atg ctc ctc aag gta aaa ctc ccc            810
Thr Pro Ser Glu Lys His Met Leu Leu Lys Val Lys Leu Pro
            165               170                   175
tgaggccgca cccatggagc ctgggcttac cctctcacct tcttcttatt aaaaatccgt     870
tttaaaaaac aaaaaaaaaa aaa                                               893
```

```
<210> 114
<211> 1475
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 55..459

<221> sig_peptide
<222> 55..120
<223> Von Heijne matrix
      score 7.2
      seq GLWLALVDGLVRS/SP

<221> polyA_signal
<222> 1444..1449

<221> polyA_site
<222> 1462..1475

<400> 114
cagttccgca gctacgtgtg ggacccgctg ctgatcctgt cgcagatcgt cctc atg            57
                                                                 Met

cag acc gtg tat tac ggc tcg ctg ggc ctg tgg ctg gcg ctg gtg gac          105
Gln Thr Val Tyr Tyr Gly Ser Leu Gly Leu Trp Leu Ala Leu Val Asp
    -20               -15               -10

ggg cta gtg cga agc agc ccc tcg ctg gac cag atg ttc gac gcc gag          153
Gly Leu Val Arg Ser Ser Pro Ser Leu Asp Gln Met Phe Asp Ala Glu
-5                  1               5                    10

atc ctg ggc ttt tcc acc cct cca ggc cgg ctc tcc atg atg tcc ttc         201
Ile Leu Gly Phe Ser Thr Pro Pro Gly Arg Leu Ser Met Met Ser Phe
            15                  20                  25

atc ttc aac gcc ctc acc tgt gcc ctg ggc ttg ctg tac ttc atc cgg         249
Ile Phe Asn Ala Leu Thr Cys Ala Leu Gly Leu Leu Tyr Phe Ile Arg
        30                  35                  40

cga gga aag cag tgt ctg gat ttc act gtc act gtc cat ttc ttt cac         297
Arg Gly Lys Gln Cys Leu Asp Phe Thr Val Thr Val His Phe Phe His
    45                  50                  55

ctc ctg ggc tgc tgg ttc tac agc tcc cgt ttc ccc tcg gcg ctg acc         345
Leu Leu Gly Cys Trp Phe Tyr Ser Ser Arg Phe Pro Ser Ala Leu Thr
60                  65                  70                  75

tgg tgg ctg gtc caa gcc gtg tgc att gca ctc atg gct gtc atc ggg         393
Trp Trp Leu Val Gln Ala Val Cys Ile Ala Leu Met Ala Val Ile Gly
                80                  85                  90

gag tac ctg tgc atg cgg acg gag ctc aag gag ata ccc ctc aac tca         441
Glu Tyr Leu Cys Met Arg Thr Glu Leu Lys Glu Ile Pro Leu Asn Ser
                95                  100                 105

gcc cct aaa tcc aat gtc tagaatcagg ccctttggac atcccgctga                 489
Ala Pro Lys Ser Asn Val
            110

cacttgggcc ccttaacacc ttgggctgct cagaccctcc agatgaggtc cagcccagat        549
ctgagaggaa ccctggaaat gtgaagtctc tgttggtgtg ggagagatag tgagggcctg        609
tcaaagaagg caggtagcag tcagcatgac agctgcaaga atgacctctg tctgttgaag        669
ccttggtatc tgagaggtca ggaagggggac ctctttgagg gtaataacat aattggaacc        729
atgccactct tgagccacaa tacctgtcac cagcctgttg ttttaagaga gaaaaaaaat        789
caaggatatc tgattggagc aaaccacttc tttagtcatc tgtcttacct ccctgggaca        849
gctgttacct ttgcagtgtt gccgaatcac agcagttacc tttgcaatgt tgccgaatca        909
cagcagttct gttggagaaa cgcttggttt ccggatccag agccacagaa agaaatgtag        969
gtgtgaagta ttaggctgct gtcagggaga ggatggcaga tggaggcatc aagcacaagg       1029
aaaatgcaca acctgtgccc tgttatacac acgttcatgt gcgcccaaga acctatgact       1089
ttcttccagt tccttctacc aggtccccat cctgctgcca gctctcaaca tagcaggcca       1149
taggacccag agaagaatcc cagtgttgct caaagtctga ccatcataaa gacactgcct       1209
gtcttctagg aatgaccagg cacccagctc ccactggact ccaatttttt ttcctgcctt       1269
atttagaatt ctttggcggg aagggtatga tgggttccca gagacaagaa gcccaacctt       1329
ctggcctggg ctgtgctgat agtgctgagg gagataggaa tttgctgcta agattttttct      1389
```

```
ttggggtgga gtttcctctg tgaggggctt gcagctatcc ttcctgtgta tacaaataca    1449
gtattttcca tgaaaaaaaa aaaaaa                                         1475
```

```
<210> 115
<211> 321
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 48..248

<221> sig_peptide
<222> 48..161
<223> Von Heijne matrix
      score 6.3
      seq LVFALVTAVCCLA/DG

<221> polyA_signal
<222> 283..288

<221> polyA_site
<222> 308..321

<400> 115
gctgagaaga gttgagggaa agtgctgctg ctgggtctgc agacgcg atg aat aac      56
                                                    Met Asn Asn
gtg cag ccg aaa ata aaa cat cgc ccc ttc tgc ttc agt gtg aaa ggc     104
Val Gln Pro Lys Ile Lys His Arg Pro Phe Cys Phe Ser Val Lys Gly
-35                     -30                 -25                 -20
cac gtg aag atg ctg cgg ctg gtg ttt gca ctt gtg aca gca gta tgc     152
His Val Lys Met Leu Arg Leu Val Phe Ala Leu Val Thr Ala Val Cys
                -15                 -10                 -5
tgt ctt gcc gac ggg gcc ctt att tac cgg aag ctt ctg ttc aat ccc     200
Cys Leu Ala Asp Gly Ala Leu Ile Tyr Arg Lys Leu Leu Phe Asn Pro
                1               5                   10
aac ggt cct tac cag aaa aag cct gtg cat gaa aaa aaa gaa gtt ttg     248
Asn Gly Pro Tyr Gln Lys Lys Pro Val His Glu Lys Lys Glu Val Leu
        15                  20                  25
tgattttata ttacttttta gtttgatact aagtattaaa catatttctg tattcttcca    308
aaaaaaaaaa aaa                                                       321
```

```
<210> 116
<211> 450
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 25..399

<221> sig_peptide
<222> 25..186
<223> Von Heijne matrix
      score 3.5
      seq SILAQVLDQSARA/RL

<400> 116
```

```
ctgctccagc gctgacgccg agcc atg gcg gac gag gag ctt gag gcg ctg      51
                          Met Ala Asp Glu Glu Leu Glu Ala Leu
                                              -50
agg aga cag agg ctg gcc gag ctg cag gcc aaa cac ggg gat cct ggt      99
Arg Arg Gln Arg Leu Ala Glu Leu Gln Ala Lys His Gly Asp Pro Gly
-45             -40             -35                     -30
gat gcg gcc caa cag gaa gca aag cac agg gaa gca gaa atg aga aac     147
Asp Ala Ala Gln Gln Glu Ala Lys His Arg Glu Ala Glu Met Arg Asn
            -25             -20                     -15
agt atc tta gcc caa gtt ctg gat cag tcg gcc cgg gcc agg tta agt     195
Ser Ile Leu Ala Gln Val Leu Asp Gln Ser Ala Arg Ala Arg Leu Ser
        -10             -5              1
aac tta gca ctt gta aag cct gaa aaa act aaa gca gta gag aat tac     243
Asn Leu Ala Leu Val Lys Pro Glu Lys Thr Lys Ala Val Glu Asn Tyr
    5               10              15
ctt ata cag atg gca aga tat gga caa cta agt gag aag gta tca gaa     291
Leu Ile Gln Met Ala Arg Tyr Gly Gln Leu Ser Glu Lys Val Ser Glu
20              25              30              35
caa ggt tta ata gaa atc ctt aaa aaa gta agc caa caa aca gaa aag     339
Gln Gly Leu Ile Glu Ile Leu Lys Lys Val Ser Gln Gln Thr Glu Lys
            40              45              50
aca aca aca gtg aaa ttc aac aga aga aaa gta atg gac tct gat gaa     387
Thr Thr Thr Val Lys Phe Asn Arg Arg Lys Val Met Asp Ser Asp Glu
            55              60              65
gat gac gat tat tgaactacaa gtgctcacag actagaactt aacggaacaa        439
Asp Asp Asp Tyr
            70
gtctaggaca g                                                       450
```

```
<210> 117
<211> 1173
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 10..1137

<221> sig_peptide
<222> 10..72
<223> Von Heijne matrix
      score 6.5
      seq LLTLLLPPPPLYT/RH

<221> polyA_signal
<222> 1144..1149

<221> polyA_site
<222> 1162..1173

<400> 117
```

```
gagctgctt atg gga cac cgc ttc ctg cgc ggc ctc tta acg ctg ctg ctg   51
            Met Gly His Arg Phe Leu Arg Gly Leu Leu Thr Leu Leu Leu
            -20                 -15                     -10
ccg ccg cca ccc ctg tat acc cgg cac cgc atg ctc ggt cca gag tcc    99
Pro Pro Pro Pro Leu Tyr Thr Arg His Arg Met Leu Gly Pro Glu Ser
        -5              1               5
gtc ccg ccc cca aaa cga tcc cgc agc aaa ctc atg gca ccg ccc cga   147
Val Pro Pro Pro Lys Arg Ser Arg Ser Lys Leu Met Ala Pro Pro Arg
10              15              20                      25
```

```
atc ggg acg cac aat ggc acc ttc cac tgc gac gag gca ctg gca tgc    195
Ile Gly Thr His Asn Gly Thr Phe His Cys Asp Glu Ala Leu Ala Cys
            30                  35                  40

gca ctg ctt cgc ctc ctg ccg gag tac cgg gat gca gag att gtg cgg    243
Ala Leu Leu Arg Leu Leu Pro Glu Tyr Arg Asp Ala Glu Ile Val Arg
            45                  50                  55

acc cgg gat ccc gaa aaa ctc gct tcc tgt gac atc gtg gtg gac gtg    291
Thr Arg Asp Pro Glu Lys Leu Ala Ser Cys Asp Ile Val Val Asp Val
            60                  65                  70

ggg ggc gag tac gac cct cgg aga cac cga tat gac cat cac cag agg    339
Gly Gly Glu Tyr Asp Pro Arg Arg His Arg Tyr Asp His His Gln Arg
            75                  80                  85

tct ttc aca gag acc atg agc tcc ctg tcc cct ggg agg ccg tgg cag    387
Ser Phe Thr Glu Thr Met Ser Ser Leu Ser Pro Gly Arg Pro Trp Gln
90                  95                  100                 105

acc aag ctg agc agt gcg gga ctc atc tat ctg cac ttc ggg cac aag    435
Thr Lys Leu Ser Ser Ala Gly Leu Ile Tyr Leu His Phe Gly His Lys
                110                 115                 120

ctg ctg gcc cag ttg ctg ggc act agt gaa gag gac agc atg gtg ggc    483
Leu Leu Ala Gln Leu Leu Gly Thr Ser Glu Glu Asp Ser Met Val Gly
                125                 130                 135

acc ctc tat gac aag atg tat gag aac ttt gtg gag gag gtg gat gct    531
Thr Leu Tyr Asp Lys Met Tyr Glu Asn Phe Val Glu Glu Val Asp Ala
                140                 145                 150

gtg gac aat ggg atc tcc cag tgg gca gag ggg gag cct cga tat gca    579
Val Asp Asn Gly Ile Ser Gln Trp Ala Glu Gly Glu Pro Arg Tyr Ala
            155                 160                 165

ctg acc act acc ctg agt gca cga gtt gct cga ctt aat cct acc tgg    627
Leu Thr Thr Thr Leu Ser Ala Arg Val Ala Arg Leu Asn Pro Thr Trp
170                 175                 180                 185

aac cac ccc gac caa gac act gag gca ggg ttc aag cgt gca atg gat    675
Asn His Pro Asp Gln Asp Thr Glu Ala Gly Phe Lys Arg Ala Met Asp
                190                 195                 200

ctg gtt caa gag gag ttt ctg cag aga tta gat ttc tac caa cac agc    723
Leu Val Gln Glu Glu Phe Leu Gln Arg Leu Asp Phe Tyr Gln His Ser
                205                 210                 215

tgg ctg cca gcc cgg gcc ttg gtg gaa gag gcc ctt gcc cag cga ttc    771
Trp Leu Pro Ala Arg Ala Leu Val Glu Glu Ala Leu Ala Gln Arg Phe
                220                 225                 230

cag gtg gac cca agt gga gag att gtg gaa ctg gcg aaa ggt gca tgt    819
Gln Val Asp Pro Ser Gly Glu Ile Val Glu Leu Ala Lys Gly Ala Cys
            235                 240                 245

ccc tgg aag gag cat ctc tac cac ctg gaa tct ggg ctg tcc cct cca    867
Pro Trp Lys Glu His Leu Tyr His Leu Glu Ser Gly Leu Ser Pro Pro
250                 255                 260                 265

gtg gcc atc ttc ttt gtt atc tac act gac cag gct gga cag tgg cga    915
Val Ala Ile Phe Phe Val Ile Tyr Thr Asp Gln Ala Gly Gln Trp Arg
                270                 275                 280

ata cag tgt gtg ccc aag gag ccc cac tca ttc caa agc cgg ctg ccc    963
Ile Gln Cys Val Pro Lys Glu Pro His Ser Phe Gln Ser Arg Leu Pro
            285                 290                 295

ctg cca gag cca tgg cgg ggt ctt cgg gac gag gcc ctg gac cag gtc   1011
Leu Pro Glu Pro Trp Arg Gly Leu Arg Asp Glu Ala Leu Asp Gln Val
            300                 305                 310

agt ggg atc cct ggc tgc atc ttc gtc cat gca agc ggc ttc att ggc   1059
Ser Gly Ile Pro Gly Cys Ile Phe Val His Ala Ser Gly Phe Ile Gly
            315                 320                 325

ggt cac cgc acc cga gag ggt gcc ttg agc atg gcc cgt gcc acc ttg   1107
Gly His Arg Thr Arg Glu Gly Ala Leu Ser Met Ala Arg Ala Thr Leu
330                 335                 340                 345

gcc cag cgc tca tac ctc cca caa atc tcc tagtctaata aaaccttcca     1157
Ala Gln Arg Ser Tyr Leu Pro Gln Ile Ser
```

```
                 350                355
                                                                    1173
tctcaaaaaa aaaaaa


<210> 118
<211> 785
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 72..704

<221> sig_peptide
<222> 72..161
<223> Von Heijne matrix
      score 13.2
      seq LLLLSTLVIPSAA/AP

<221> polyA_signal
<222> 772..777


<400> 118
cggaatccgg gagtccggtg acccgggctg tggtctagca taaaggcgga gcccagaaga     60
aggggcgggg t atg gga gaa gcc tcc cca cct gcc ccc gca agg cgg cat    110
            Met Gly Glu Ala Ser Pro Pro Ala Pro Ala Arg Arg His
            -30               -25               -20
ctg ctg gtc ctg ctg ctg ctc ctc tct acc ctg gtg atc ccc tcc gct    158
Leu Leu Val Leu Leu Leu Leu Leu Ser Thr Leu Val Ile Pro Ser Ala
        -15               -10                -5
gca gct cct atc cat gat gct gac gcc caa gag agc tcc ttg ggt ctc    206
Ala Ala Pro Ile His Asp Ala Asp Ala Gln Glu Ser Ser Leu Gly Leu
1          5                    10                 15
aca ggc ctc cag agc cta ctc caa ggc ttc agc cga ctt ttc ctg aaa    254
Thr Gly Leu Gln Ser Leu Leu Gln Gly Phe Ser Arg Leu Phe Leu Lys
                20                 25                 30
ggt aac ctg ctt cgg ggc ata gac agc tta ttc tct gcc ccc atg gac    302
Gly Asn Leu Leu Arg Gly Ile Asp Ser Leu Phe Ser Ala Pro Met Asp
            35                 40                 45
ttc cgg ggc ctc cct ggg aac tac cac aaa gag gag aac cag gag cac    350
Phe Arg Gly Leu Pro Gly Asn Tyr His Lys Glu Glu Asn Gln Glu His
        50                 55                 60
cag ctg ggg aac aac acc ctc tcc agc cac ctc cag atc gac aag gta    398
Gln Leu Gly Asn Asn Thr Leu Ser Ser His Leu Gln Ile Asp Lys Val
65                 70                 75
ccc agg atg gag gag aag gag gcc ctg gta ccc atc cag aag gcc acg    446
Pro Arg Met Glu Glu Lys Glu Ala Leu Val Pro Ile Gln Lys Ala Thr
80                 85                 90                 95
gac agc ttc cac aca gaa ctc cat ccc cgg gtg gcc ttc tgg atc att    494
Asp Ser Phe His Thr Glu Leu His Pro Arg Val Ala Phe Trp Ile Ile
                100                105                110
aag ctg cca cgg cgg agg tcc cac cag gat gcc ctg gag ggc ggc cac    542
Lys Leu Pro Arg Arg Arg Ser His Gln Asp Ala Leu Glu Gly Gly His
            115                120                125
tgg ctc agc gag aag cga cac cgc ctg cag gcc atc cgg gat gga ctc    590
Trp Leu Ser Glu Lys Arg His Arg Leu Gln Ala Ile Arg Asp Gly Leu
            130                135                140
cgc aag ggg acc cac aag gac gtc cta gaa gag ggg acc gag agc tcc    638
Arg Lys Gly Thr His Lys Asp Val Leu Glu Glu Gly Thr Glu Ser Ser
145                150                155
tcc cac tcc agg ctg tcc ccc cga aag acc cac tta ctg tac atc ctc    686
```

```
Ser His Ser Arg Leu Ser Pro Arg Lys Thr His Leu Leu Tyr Ile Leu
160              165              170              175
agg ccc tct cgg cag ctg taggggtggg gaccggggag cacctgcctg            734
Arg Pro Ser Arg Gln Leu
              180
tagcccccat cagaccctgc cccaagcacc atatggaaat aaagttcttt c            785
```

```
<210> 119
<211> 559
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 44..505

<221> sig_peptide
<222> 44..223
<223> Von Heijne matrix
      score 4
      seq LVRRTLLVAALRA/WM

<400> 119
agcaaccaga gggagatgat cacctgaacc actgctccaa acc atg ggc agt aaa     55
                                                 Met Gly Ser Lys
                                                 -60
tgc tgt aaa ggt ggt cca gat gaa gat gca gta gaa aga cag agg cgg    103
Cys Cys Lys Gly Gly Pro Asp Glu Asp Ala Val Glu Arg Gln Arg Arg
    -55              -50              -45
cag aag ttg ctt ctt gca caa ctg cat cac aga aaa agg gtg aag gca    151
Gln Lys Leu Leu Leu Ala Gln Leu His His Arg Lys Arg Val Lys Ala
-40              -35              -30              -25
gct ggg cag atc cag gcc tgg tgg cgt ggg gtc ctg gtg cgc agg acc    199
Ala Gly Gln Ile Gln Ala Trp Trp Arg Gly Val Leu Val Arg Arg Thr
              -20              -15              -10
ctg ctg gtt gct gcc ctc agg gcc tgg atg att cag tgc tgg tgg agg    247
Leu Leu Val Ala Ala Leu Arg Ala Trp Met Ile Gln Cys Trp Trp Arg
              -5               1                5
acg ttg gtg cag aga cgg atc cgt cag cgg cgg cag gcc ctg ttg agg    295
Thr Leu Val Gln Arg Arg Ile Arg Gln Arg Arg Gln Ala Leu Leu Arg
              10               15               20
gtc tac gtc atc cag gag cag gcg acg gtc aag ctc cag tcc tgc atc    343
Val Tyr Val Ile Gln Glu Gln Ala Thr Val Lys Leu Gln Ser Cys Ile
25               30               35               40
cgc atg tgg cag tgc cgg caa tgt tac cgc caa atg tgc aat gct ctc    391
Arg Met Trp Gln Cys Arg Gln Cys Tyr Arg Gln Met Cys Asn Ala Leu
              45               50               55
tgc ttg ttc cag gtc cca gag agc agc ctt gcc ttc cag act gat ggc    439
Cys Leu Phe Gln Val Pro Glu Ser Ser Leu Ala Phe Gln Thr Asp Gly
              60               65               70
ttt tta cag gtc caa tat gca atc cct tca aag cag cca gag ttc cac    487
Phe Leu Gln Val Gln Tyr Ala Ile Pro Ser Lys Gln Pro Glu Phe His
              75               80               85
att gaa atc cta tca atc tgaaaggcct ggggcatgga gaacaggctg            535
Ile Glu Ile Leu Ser Ile
              90
cactaccctа ataaatgtct gacc                                          559
```

```
<210> 120
<211> 770
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 25..393

<221> sig_peptide
<222> 25..150
<223> Von Heijne matrix
      score 4.6
      seq LDPAVSLSAPAFA/SA

<221> polyA_signal
<222> 734..739

<221> polyA_site
<222> 757..770

<400> 120
cgcagaaagg agagacacac atac atg aaa gga gga gct ttc tcc aat ctt        51
                          Met Lys Gly Gly Ala Phe Ser Asn Leu
                          -40                         -35
aat gat tcc cag ctc tca gcc tcg ttt ctg caa ccc agc ctg caa gca      99
Asn Asp Ser Gln Leu Ser Ala Ser Phe Leu Gln Pro Ser Leu Gln Ala
        -30                 -25                 -20
aac tgt cct gct ttg gac cct gct gtg tca ctc tcc gca cca gcc ttt     147
Asn Cys Pro Ala Leu Asp Pro Ala Val Ser Leu Ser Ala Pro Ala Phe
        -15                 -10                  -5
gcc tct gct ctt cgc tct atg aag tcc tcc cag gct gca cgg aag gac     195
Ala Ser Ala Leu Arg Ser Met Lys Ser Ser Gln Ala Ala Arg Lys Asp
 1               5                   10                  15
gac ttt ctc agg tct ctt agt gat gga gac tca ggg aca tca gaa cac     243
Asp Phe Leu Arg Ser Leu Ser Asp Gly Asp Ser Gly Thr Ser Glu His
            20                  25                  30
atc tca gcg gtg gtg act agc cct cgg att tcc tgc cat ggt gct gcc     291
Ile Ser Ala Val Val Thr Ser Pro Arg Ile Ser Cys His Gly Ala Ala
            35                  40                  45
att ccc acc gcc cgt gcc ctc tgc cta ggc tgt tcc tgc tgc acc gaa     339
Ile Pro Thr Ala Arg Ala Leu Cys Leu Gly Cys Ser Cys Cys Thr Glu
         50                  55                  60
cgc ctc ctc ctg cca ccg ccc tcc ctc ctt tct tta gaa gcc cct gcc     387
Arg Leu Leu Leu Pro Pro Pro Ser Leu Leu Ser Leu Glu Ala Pro Ala
         65                  70                  75
agc acc tgagctctct gctgattgct gttcctccca gtctgtggaa gctttgccca      443
Ser Thr
80
tatgctttcc ttaaaagggt tctgggcagg gcaggcgccc ccatttctca gggatcccct    503
ccaggacaac gccttttcct tgtgtcttca gctctcctta ccagatatct atatatttgt    563
atatattcag tttcaccaac aatgcatcaa gtactttttt ttttaagtaa agaaccgcag    623
tcatcgaact ggagccccat tgattccctc cccctcgcct ccccaaatct ggcacctgcc    683
caaggtatcc tcagaaccat ttggggtgtc ctttggcatt ggataataga aataaaattt    743
tacctctttc tacaaaaaaa aaaaaac                                         770


<210> 121
<211> 1213
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 58..1095

<221> sig_peptide
<222> 58..114
<223> Von Heijne matrix
      score 5.4
      seq LSHLLPSLRQVIQ/EP

<221> polyA_site
<222> 1202..1213

<400> 121
cctggctttg cctttgccct gctgtgtgat cttagctccc tgcccaggcc cacagcc           57
atg gcc atg gcc cag aaa ctc agc cac ctc ctg ccg agt ctg cgg cag          105
Met Ala Met Ala Gln Lys Leu Ser His Leu Leu Pro Ser Leu Arg Gln
            -15                 -10                 -5
gtc atc cag gag cct cag cta tct ctg cag cca gag cct gtc ttc acg          153
Val Ile Gln Glu Pro Gln Leu Ser Leu Gln Pro Glu Pro Val Phe Thr
  1                   5                  10
gtg gat cga gct gag gtg ccg ccg ctc ttc tgg aag ccg tac atc tat          201
Val Asp Arg Ala Glu Val Pro Pro Leu Phe Trp Lys Pro Tyr Ile Tyr
 15                  20                  25
gcg ggc tac cgg ccg ctg cat cag acc tgg cgc ttc tat ttc cgc acg          249
Ala Gly Tyr Arg Pro Leu His Gln Thr Trp Arg Phe Tyr Phe Arg Thr
30                  35                  40                  45
ctg ttc cag cag cac aac gag gcc gtg aat gtc tgg acc cac ctg ctg          297
Leu Phe Gln Gln His Asn Glu Ala Val Asn Val Trp Thr His Leu Leu
                50                  55                  60
gcg gcc ctg gta ctg ctg ctg cgg ctg gcc ctc ttt gtg gag acc gtg          345
Ala Ala Leu Val Leu Leu Leu Arg Leu Ala Leu Phe Val Glu Thr Val
            65                  70                  75
gac ttc tgg gga gac cca cac gcc ctg ccc ctc ttc atc att gtc ctt          393
Asp Phe Trp Gly Asp Pro His Ala Leu Pro Leu Phe Ile Ile Val Leu
            80                  85                  90
gcc tct ttc acc tac ctc tcc ctc agt gcc ttg gct cac ctc ctg cag          441
Ala Ser Phe Thr Tyr Leu Ser Leu Ser Ala Leu Ala His Leu Leu Gln
95                  100                 105
gcc aag tct gag ttc tgg cat tac agc ttc ttc ttc ctg gac tat gtg          489
Ala Lys Ser Glu Phe Trp His Tyr Ser Phe Phe Phe Leu Asp Tyr Val
110                 115                 120                 125
ggg gtg gcc gtg tac cag ttt ggc agt gcc ttg gca cac ttc tac tat          537
Gly Val Ala Val Tyr Gln Phe Gly Ser Ala Leu Ala His Phe Tyr Tyr
                130                 135                 140
gct atc gag ccc gcc tgg cat gcc cag gtg cag gct gtt ttt ctg ccc          585
Ala Ile Glu Pro Ala Trp His Ala Gln Val Gln Ala Val Phe Leu Pro
                145                 150                 155
atg gct gcc ttt ctc gcc tgg ctt tcc tgc att ggc tcc tgc tat aac          633
Met Ala Ala Phe Leu Ala Trp Leu Ser Cys Ile Gly Ser Cys Tyr Asn
            160                 165                 170
aag tac atc cag aaa cca ggc ctg ctg ggc cgc aca tgc cag gag gtg          681
Lys Tyr Ile Gln Lys Pro Gly Leu Leu Gly Arg Thr Cys Gln Glu Val
            175                 180                 185
ccc tcc gtc ctg gcc tac gca ctg gac att agt cct gtg gtg cat cgt          729
Pro Ser Val Leu Ala Tyr Ala Leu Asp Ile Ser Pro Val Val His Arg
190                 195                 200                 205
atc ttc gtg tcc tcc gac ccc acc acg gat gat cca gct ctt ctc tac          777
Ile Phe Val Ser Ser Asp Pro Thr Thr Asp Asp Pro Ala Leu Leu Tyr
                210                 215                 220
cac aag tgc cag gtg gtc ttc ttt ctg ctg gct gct gcc ttc ttc tct          825
```

```
His Lys Cys Gln Val Val Phe Phe Leu Leu Ala Ala Ala Phe Phe Ser
            225                 230                 235
acc ttc atg ccc gag cgc tgg ttc cct ggc agc tgc cat gtc ttc ggg      873
Thr Phe Met Pro Glu Arg Trp Phe Pro Gly Ser Cys His Val Phe Gly
            240                 245                 250
cag ggc cac caa ctt ttc cat atc ttc ttg gtg ctg tgc acg ctg gct      921
Gln Gly His Gln Leu Phe His Ile Phe Leu Val Leu Cys Thr Leu Ala
            255                 260                 265
cag ctg gag gct gtg gca ctg gac tat gag gcc cga cgg ccc atc tat      969
Gln Leu Glu Ala Val Ala Leu Asp Tyr Glu Ala Arg Arg Pro Ile Tyr
270                 275                 280                 285
gag cct ctg cac acg cac tgg cct cac aac ttt tct ggc ctc ttc ctg     1017
Glu Pro Leu His Thr His Trp Pro His Asn Phe Ser Gly Leu Phe Leu
                290                 295                 300
ctc acg gtg ggc agc agc atc ctc act gca ttc ctc ctg agc cag ctg     1065
Leu Thr Val Gly Ser Ser Ile Leu Thr Ala Phe Leu Leu Ser Gln Leu
            305                 310                 315
gta cag cgc aaa ctt gat cag aag acc aag tgaaggggga tggcatctgg      1115
Val Gln Arg Lys Leu Asp Gln Lys Thr Lys
            320                 325
tagggaggga ggtatagttg ggggacaggg gtctgggttt ggctccaagt gggaacaagg    1175
cctggtaaag ttgtttgtgt ctggccaaaa aaaaaaaa                            1213
```

```
<210> 122
<211> 1318
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 31..660

<221> sig_peptide
<222> 31..90
<223> Von Heijne matrix
      score 5.4
      seq AFVIACVLSLIST/IY

<221> polyA_signal
<222> 1288..1293

<221> polyA_site
<222> 1307..1318

<400> 122
ggaggatggg cgagcagtct gaatgccaga atg gat aac cgt ttt gct aca gca       54
                                    Met Asp Asn Arg Phe Ala Thr Ala
                                    -20                 -15
ttt gta att gct tgt gtg ctt agc ctc att tcc acc atc tac atg gca      102
Phe Val Ile Ala Cys Val Leu Ser Leu Ile Ser Thr Ile Tyr Met Ala
            -10                 -5                  1
gct tcc att ggc aca gac ttc tgg tat gag tat cga agt cca gtt caa      150
Ala Ser Ile Gly Thr Asp Phe Trp Tyr Glu Tyr Arg Ser Pro Val Gln
5                   10                  15                  20
gaa aat tcc agt gat ttg aat aaa agc atc tgg gat gaa ttc att agt      198
Glu Asn Ser Ser Asp Leu Asn Lys Ser Ile Trp Asp Glu Phe Ile Ser
            25                  30                  35
gat gag gca gat gaa aag act tat aat gat gca ctt ttt cga tac aat      246
Asp Glu Ala Asp Glu Lys Thr Tyr Asn Asp Ala Leu Phe Arg Tyr Asn
            40                  45                  50
```

```
ggc aca gtg gga ttg tgg aga cgg tgt atc acc ata ccc aaa aac atg        294
Gly Thr Val Gly Leu Trp Arg Arg Cys Ile Thr Ile Pro Lys Asn Met
        55                      60                      65
cat tgg tat agc cca cca gaa agg aca gag tca ttt gat gtg gtc aca        342
His Trp Tyr Ser Pro Pro Glu Arg Thr Glu Ser Phe Asp Val Val Thr
    70                      75                      80
aaa tgt gtg agt ttc aca cta act gag cag ttc atg gag aaa ttt gtt        390
Lys Cys Val Ser Phe Thr Leu Thr Glu Gln Phe Met Glu Lys Phe Val
85                      90                      95                     100
gat ccc gga aac cac aat agc ggg att gat ctc ctt agg acc tat ctt        438
Asp Pro Gly Asn His Asn Ser Gly Ile Asp Leu Leu Arg Thr Tyr Leu
            105                     110                     115
tgg cgt tgc cag ttc ctt tta cct ttt gtg agt tta ggt ttg atg tgc        486
Trp Arg Cys Gln Phe Leu Leu Pro Phe Val Ser Leu Gly Leu Met Cys
            120                     125                     130
ttt ggg gct ttg atc gga ctt tgt gct tgc att tgc cga agc tta tat        534
Phe Gly Ala Leu Ile Gly Leu Cys Ala Cys Ile Cys Arg Ser Leu Tyr
            135                     140                     145
ccc acc att gcc acg ggc att ctc cat ctc ctt gca gtg aca aag gag        582
Pro Thr Ile Ala Thr Gly Ile Leu His Leu Leu Ala Val Thr Lys Glu
            150                     155                     160
agc atg ctt cca gct gga gct gag tcc aag cac aca gcc act cct gca        630
Ser Met Leu Pro Ala Gly Ala Glu Ser Lys His Thr Ala Thr Pro Ala
165                     170                     175                     180
cac gca tgc gtg caa aca ggg aag ccc aag taggagaaga ggaaagaggt        680
His Ala Cys Val Gln Thr Gly Lys Pro Lys
            185                     190
tgtagggatt tgggaagaac cttgattatt ccctggagga aaagacaaat ctacttccct     740
gaaatcaccc tcgaatctac ttccaccctc agaacttaaa atgaactgca tccttttttt     800
catcttcttt tcttctccag tgaatatgat ctccaaaccc ttatttttc tttgaactgt      860
aaaattcca ctcatggacg atgcaaccaa cagatgcaat ctctgagaag atgaaaattg      920
ggacctctta ttataaaatt gacctagctg gactcaggaa accagggaag aagtcaatgc     980
aggcatttaa aatgtaaagt tttttctggt taaatctatt tattttctt gtaggttgag     1040
tatttcttcc cagttttttct gctctggtgt ataacaaaca ggtcaaaatt tcccatcttt    1100
cctcctgata gtagttgaat cctaccttgc atacttaatg catagtgaaa tggcatctag     1160
cagaaataca cacccccaaa acacaccacc atttcattag gtgcccaaaa aattctgtat     1220
ttagcttatt tatttattgt tattttgct ttttcttaac ccactatata ttgactgcaa     1280
acgaattaat aaaattatccc ttctggaaaa aaaaaaaa                           1318
                                                                      1318
```

<210> 123
<211> 853
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 31..582

<221> sig_peptide
<222> 31..90
<223> Von Heijne matrix
      score 5.4
      seq AFVIACVLSLIST/IY

<221> polyA_signal
<222> 816..821

<221> polyA_site
<222> 840..853

```
<400> 123
ggaggatggg cgagcagtct gaatgccaga atg gat aac cgt ttt gct aca gca    54
                                  Met Asp Asn Arg Phe Ala Thr Ala
                                  -20                   -15

ttt gta att gct tgt gtg ctt agc ctc att tcc acc atc tac atg gca    102
Phe Val Ile Ala Cys Val Leu Ser Leu Ile Ser Thr Ile Tyr Met Ala
      -10               -5                   1

gcc tcc att ggc aca gac ttc tgg tat gaa tat cga agt cca gtt caa    150
Ala Ser Ile Gly Thr Asp Phe Trp Tyr Glu Tyr Arg Ser Pro Val Gln
5                   10                  15                  20

gaa aat tcc agt gat ttg aat aaa agc atc tgg gat gaa ttc att agt    198
Glu Asn Ser Ser Asp Leu Asn Lys Ser Ile Trp Asp Glu Phe Ile Ser
                25                  30                  35

gat gaa gca gat gaa aag act tat aat gat gca cct ttt cga tac aat    246
Asp Glu Ala Asp Glu Lys Thr Tyr Asn Asp Ala Pro Phe Arg Tyr Asn
            40                  45                  50

ggc aca gtg gga ttg tgg aga cgg tgt atc acc ata ccc aaa aac atg    294
Gly Thr Val Gly Leu Trp Arg Arg Cys Ile Thr Ile Pro Lys Asn Met
        55                  60                  65

cat tgg tat agc cca cca gaa agg aca gag tca ttt gat gtg gtc aca    342
His Trp Tyr Ser Pro Pro Glu Arg Thr Glu Ser Phe Asp Val Val Thr
    70                  75                  80

aaa tgt gtg agt ttc aca cta act gag cag ttc atg gag aaa ttt gtt    390
Lys Cys Val Ser Phe Thr Leu Thr Glu Gln Phe Met Glu Lys Phe Val
85                  90                  95                  100

gat ccc gga aac cac aat agc ggg att gat ctc ctt agg acc tat ctt    438
Asp Pro Gly Asn His Asn Ser Gly Ile Asp Leu Leu Arg Thr Tyr Leu
                105                 110                 115

tgg cgt tgc cag ttc ctt tta cct ttt gtg agt tta ggt ttg atg tgc    486
Trp Arg Cys Gln Phe Leu Leu Pro Phe Val Ser Leu Gly Leu Met Cys
            120                 125                 130

ttt ggg gct ttg atc gga ctt tgt gct tgc att tgc cga agc tta tat    534
Phe Gly Ala Leu Ile Gly Leu Cys Ala Cys Ile Cys Arg Ser Leu Tyr
        135                 140                 145

ccc acc att gcc acg ggc att ctc cat ctc ctt gca gat acc atg ctg    582
Pro Thr Ile Ala Thr Gly Ile Leu His Leu Leu Ala Asp Thr Met Leu
        150                 155                 160

tgaagtccag gccacatgga ggtgtcctgt gtagatgctc cagctgaaat cccaagctaa   642
gctcccaact gacagccaac atcatttcca gccatgtgtg ggagccatcc tggatgtcca   702
gccttaacaa gccttcagag gacttcagcc acagctatta tcttactaca tccttgtgag   762
actctaataa agaaccaact agctgagccc aatcaaccta tggaactgat agaaataaaa   822
tgaattgttg ttttgcgaaa aaaaaaaaaa a                                  853
```

```
<210> 124
<211> 826
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 15..695

<221> sig_peptide
<222> 15..80
<223> Von Heijne matrix
      score 8.5
      seq AALLLGLMMVVTG/DE

<221> polyA_signal
<222> 795..800
```

```
<221> polyA_site
<222> 814..826

<400> 124
aaccagaggt gccc atg ggt tgg aca atg agg ctg gtc aca gca gca ctg          50
             Met Gly Trp Thr Met Arg Leu Val Thr Ala Ala Leu
             -20                       -15

tta ctg ggt ctc atg atg gtg gtc act gga gac gag gat gag aac agc          98
Leu Leu Gly Leu Met Met Val Val Thr Gly Asp Glu Asp Glu Asn Ser
-10             -5                 1                 5

ccg tgt gcc cat gag gcc ctc ctg gac gag gac acc ctc ttt tgc cag         146
Pro Cys Ala His Glu Ala Leu Leu Asp Glu Asp Thr Leu Phe Cys Gln
            10              15              20

ggc ctt gaa gtt ttc tac cca gag ttg ggg aac att ggc tgc aag gtt         194
Gly Leu Glu Val Phe Tyr Pro Glu Leu Gly Asn Ile Gly Cys Lys Val
         25              30              35

gtt cct gat tgt aac aac tac aga cag aag atc acc tcc tgg atg gag         242
Val Pro Asp Cys Asn Asn Tyr Arg Gln Lys Ile Thr Ser Trp Met Glu
    40              45              50

ccg ata gtc aag ttc ccg ggg gcc gtg gac ggc gca acc tat atc ctg         290
Pro Ile Val Lys Phe Pro Gly Ala Val Asp Gly Ala Thr Tyr Ile Leu
55              60              65              70

gtg atg gtg gat cca gat gcc cct agc aga gca gaa ccc aga cag aga         338
Val Met Val Asp Pro Asp Ala Pro Ser Arg Ala Glu Pro Arg Gln Arg
            75              80              85

ttc tgg aga cat tgg ctg gta aca gat atc aag ggc gcc gac ctg aag         386
Phe Trp Arg His Trp Leu Val Thr Asp Ile Lys Gly Ala Asp Leu Lys
         90              95             100

aaa ggg aag att cag ggc cag gag tta tca gcc tac cag gct ccc tcc         434
Lys Gly Lys Ile Gln Gly Gln Glu Leu Ser Ala Tyr Gln Ala Pro Ser
        105             110             115

cca ccg gca cac agt ggc ttc cat cgc tac cag ttc ttt gtc tat ctt         482
Pro Pro Ala His Ser Gly Phe His Arg Tyr Gln Phe Phe Val Tyr Leu
    120             125             130

cag gaa gga aag gtc atc tct ctc ctt ccc aag gaa aac aaa act cga         530
Gln Glu Gly Lys Val Ile Ser Leu Leu Pro Lys Glu Asn Lys Thr Arg
135             140             145             150

ggc tct tgg aaa atg gac aga ttt ctg aac cgt ttc cac ctg ggc gaa         578
Gly Ser Trp Lys Met Asp Arg Phe Leu Asn Arg Phe His Leu Gly Glu
            155             160             165

cct gaa gca agc acc cag ttc atg acc cag aac tac cag gac tca cca         626
Pro Glu Ala Ser Thr Gln Phe Met Thr Gln Asn Tyr Gln Asp Ser Pro
            170             175             180

acc ctc cag gct ccc aga gaa agg gcc agc gag ccc aag cac aaa aac         674
Thr Leu Gln Ala Pro Arg Glu Arg Ala Ser Glu Pro Lys His Lys Asn
            185             190             195

cag gcg gag ata gct gcc tgc tagatagccg gctttgccat ccgggcatgt           725
Gln Ala Glu Ile Ala Ala Cys
            200             205

ggccacactg cccaccaccg acgatgtggg tatggaaccc cctctggata cagaacccct      785
tctttccaa ataaaaaaaa aatcatccaa aaaaaaaaaa a                          826



<210> 125
<211> 571
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
```

```
<222> 74..295

<221> sig_peptide
<222> 74..196
<223> Von Heijne matrix
      score 5.4
      seq RLLYIGFLGYCSG/LI

<221> polyA_signal
<222> 545..550

<221> polyA_site
<222> 561..571

<400> 125
cgggtagtgg tcgtcgtggt tttccttgta gttcgtggtc tgagaccagg cctcaagtgg      60
aaacggcgtc acc atg atc gca cgg cgg aac cca gta ccc tta cgg ttt        109
            Met Ile Ala Arg Arg Asn Pro Val Pro Leu Arg Phe
              -40             -35             -30
ctg ccg gat gag gcc cgg agc ctg ccc ccg ccc aag ctg acc gac ccg       157
Leu Pro Asp Glu Ala Arg Ser Leu Pro Pro Pro Lys Leu Thr Asp Pro
          -25             -20             -15
cgg ctc ctc tac atc ggc ttc ttg ggc tac tgc tcc ggc ctg att gat       205
Arg Leu Leu Tyr Ile Gly Phe Leu Gly Tyr Cys Ser Gly Leu Ile Asp
          -10             -5              1
aac ctg atc cgg cgg agg ccg atc gcg acg gct ggt ttg cat cgc cag       253
Asn Leu Ile Arg Arg Arg Pro Ile Ala Thr Ala Gly Leu His Arg Gln
    5               10              15
ctt cta tat att acg gcc ttt ttt ttg ctg gat att atc ttg               295
Leu Leu Tyr Ile Thr Ala Phe Phe Leu Leu Asp Ile Ile Leu
20              25              30
taaaacgtga agactacctg tatgctgtga gggaccgtga aatgtttgga tatatgaaat      355
tacatccaga ggattttcct gaagaagata agaaaacata tggtgaaatt tttgaaaaat      415
tccatccaat acgttgaagt cttcaaaatg cttgctccag tttcactgat acctgctgtt      475
cctgaatttg atggaacatg tttcttatga cagttgaagc ttatgctaat ctgtatgttg      535
acaccttgta attaaaatac gtaccaaaaa aaaaaa                                571
```

```
<210> 126
<211> 659
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 440..658

<221> polyA_signal
<222> 601..606

<400> 126
cgccttacga gctgggaggt ggtgcctctc acccagctaa ttgctctcta gcccttggcc      60
ttcacaggtg ttggtgcctg ccgtgaacgc attctgacct gggccgtatc tgtctcccaa     120
gactttgtgc ctatggttgg ggacagagtg aggtcgttgc cttgacgacg acagcatgcg     180
gcccgtggtc ctcctaagtg tgagcttgcg gcggaccgag gcccacctgc ctccctgcct     240
gcttcgccca ggactcgtga ctgcgtccgc agaagaaatc acaacagcgc tggaattgct     300
agtttgctag gcagcatctt ttggacctgc gaaccatatg catttcacct caaatctgtt     360
tccaagttga aaacctttgg gtctttctat gcgaacggat tgaagaaacg caaaaagttt     420
ctacggactt taaattaaaa atg gaa aaa tat gaa aac ctg ggt ttg gtt gga     472
                       Met Glu Lys Tyr Glu Asn Leu Gly Leu Val Gly
                        1               5               10
```

```
gaa ggg agt tat gga atg gtg atg aag tgt agg aat aaa gat act gga      520
Glu Gly Ser Tyr Gly Met Val Met Lys Cys Arg Asn Lys Asp Thr Gly
         15              20                  25
aga att gtg gcc ata aag aag ttc tta gaa agt gac gat gac aaa atg      568
Arg Ile Val Ala Ile Lys Lys Phe Leu Glu Ser Asp Asp Asp Lys Met
     30              35                  40
gtt aaa aag att gca atg cga gaa gtc aag tta cta aag caa ctt agg      616
Val Lys Lys Ile Ala Met Arg Glu Val Lys Leu Leu Lys Gln Leu Arg
     45              50                  55
cat gaa aac ttg gtg aat ctc ttg gaa gtg tgt aaa aaa aaa a           659
His Glu Asn Leu Val Asn Leu Leu Glu Val Cys Lys Lys Lys
60                  65                  70
```

```
<210> 127
<211> 301
<212> DN.ı
<213> Homo sapiens

<220>
<221> CDS
<222> 38..283

<221> sig_peptide
<222> 38..85
<223> Von Heijne matrix
      score 4.1
      seq LLPATSLAGPVLS/TL

<221> polyA_signal
<222> 257..262

<400> 127
cacctgaatc ccaggaaccc tcaatgaggt cttcaag atg aag aga ctg ctg cca      55
                                             Met Lys Arg Leu Leu Pro
                                                 -15
gct acc agc ctg gct ggc cct gtc ctg tcc acc ctc att gcc cca act      103
Ala Thr Ser Leu Ala Gly Pro Val Leu Ser Thr Leu Ile Ala Pro Thr
-10              -5                  1                  5
ccc atg ttg ttt tgt gaa gat aaa agc tgg gat ctt ttt ctt ttt ttt      151
Pro Met Leu Phe Cys Glu Asp Lys Ser Trp Asp Leu Phe Leu Phe Phe
             10                  15                  20
aag tct cac aag aca tgg ggc atc tcc aca aat tta agt tcc tgt cca      199
Lys Ser His Lys Thr Trp Gly Ile Ser Thr Asn Leu Ser Ser Cys Pro
     25                  30                  35
ttt gga aat ttg ttt cta tgt gta cag ttt gtc aga gaa aaa caa agt      247
Phe Gly Asn Leu Phe Leu Cys Val Gln Phe Val Arg Glu Lys Gln Ser
     40                  45                  50
ttt tgt atg aat aca gaa tgt gat tta cgc aag aat tgacaaaaaa         293
Phe Cys Met Asn Thr Glu Cys Asp Leu Arg Lys Asn
55                  60                  65
aaaaaaaa                                                            301
```

```
<210> 128
<211> 477
<212> DNA
<213> Homo sapiens

<220>
```

```
<221> CDS
<222> 121..477

<221> sig_peptide
<222> 121..288
<223> Von Heijne matrix
      score 3.5
      seq SSCADSFVSSSSS/QP


<400> 128
cctcggagca ggcggagtaa agggacttga gcgagccagt tgccggatta ttctatttcc    60
cctccctctc tcccgccccg tatctctttt caccottctc ccaccctcgc tcgcgtagcc.  120
atg gcg gag ccg tcg gcg gcc act cag tcc cat tcc atc tcc tcg tcg    168
Met Ala Glu Pro Ser Ala Ala Thr Gln Ser His Ser Ile Ser Ser Ser
    -55             -50             -45
tcc ttc gga gcc gag ccg tcc gcg ccc ggc ggc ggc ggg agc cca gga    216
Ser Phe Gly Ala Glu Pro Ser Ala Pro Gly Gly Gly Gly Ser Pro Gly
-40             -35             -30             -25
gcc tgc ccc gcc ctg ggg acg aag agc tgc agc tcc tcc tgt gcg gat    264
Ala Cys Pro Ala Leu Gly Thr Lys Ser Cys Ser Ser Ser Cys Ala Asp
                -20             -15             -10
tcc ttt gtt tct tcc tct tcc tct cag cct gta tct cta ttt tcg acc    312
Ser Phe Val Ser Ser Ser Ser Ser Gln Pro Val Ser Leu Phe Ser Thr
                -5              1               5
tca caa gag gga ttg agc tct ctt tgc tct gat gag cca tct tca gaa    360
Ser Gln Glu Gly Leu Ser Ser Leu Cys Ser Asp Glu Pro Ser Ser Glu
    10              15              20
att atg act tct tcc ttt ctt tca tct tct gaa ata cat aac act ggc    408
Ile Met Thr Ser Ser Phe Leu Ser Ser Ser Glu Ile His Asn Thr Gly
25              30              35              40
ctt aca ata cta cat gga gaa aaa agc cat gtg tta ggg agc cag cct    456
Leu Thr Ile Leu His Gly Glu Lys Ser His Val Leu Gly Ser Gln Pro
                45              50              55
att tta gcc aaa aaa aaa aaa                                         477
Ile Leu Ala Lys Lys Lys Lys
                60




<210> 129
<211> 323
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 2..163

<221> polyA_signal
<222> 292..297

<221> polyA_site
<222> 310..323


<400> 129
a gct ttc gtg tgg gag cca gct atg gtg cgg atc aat gcg ctg aca gca    49
  Ala Phe Val Trp Glu Pro Ala Met Val Arg Ile Asn Ala Leu Thr Ala
  1             5               10              15
gcc tct gag gct gcg tgc ctg atc gtg tct gta gat gaa acc atc aag    97
Ala Ser Glu Ala Ala Cys Leu Ile Val Ser Val Asp Glu Thr Ile Lys
            20  .            25  '            30
aac ccc cgc tcg act gtg gat gct ccc aca gca gca ggc cgg ggc cgt   145
```

```
Asn Pro Arg Ser Thr Val Asp Ala Pro Thr Ala Ala Gly Arg Gly Arg
        35                  40                  45
ggt cgt ggc cgc ccc cac tgagaggcac cccacccatc acatggctgg          193
Gly Arg Gly Arg Pro His
        50
ctggctgctg ggtgcactta ccctccttgg cttggttact tcattttaca aggaaggggt  253
agtaattggc ccactctctt cttactggag gctatttaaa taaaatgtaa gacttcaaaa  313
aaaaaaaaaa                                                          323
```

```
<210> 130
<211> 1392
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 46..675

<221> sig_peptide
<222> 46..87
<223> Von Heijne matrix
      score 5.3
      seq LTLLGLSFILAGL/IV

<221> polyA_signal
<222> 1364..1369

<221> polyA_site
<222> 1383..1392

<400> 130
ctccgagttg ccacccagga aaaagagggc tcctctggga gatgt atg ctt act ctc   57
                                                    Met Leu Thr Leu
tta ggc ctt tca ttc atc ttg gca gga ctt att gtt ggt gga gcc tgc   105
Leu Gly Leu Ser Phe Ile Leu Ala Gly Leu Ile Val Gly Gly Ala Cys
-10             -5                  1                   5
att tac aag tac ttc atg ccc aag agc acc att tac cgt gga gag atg   153
Ile Tyr Lys Tyr Phe Met Pro Lys Ser Thr Ile Tyr Arg Gly Glu Met
            10                  15                  20
tgc ttt ttt gat tct gag gat cct gca aat tcc ctt cgt gga gga gag   201
Cys Phe Phe Asp Ser Glu Asp Pro Ala Asn Ser Leu Arg Gly Gly Glu
        25                  30                  35
cct aac ttc ctg cct gtg act gag gag gct gac att cgt gag gat gac   249
Pro Asn Phe Leu Pro Val Thr Glu Glu Ala Asp Ile Arg Glu Asp Asp
        40                  45                  50
aac att gca atc att gat gtg cct gtc ccc agt ttc tct gat agt gac   297
Asn Ile Ala Ile Ile Asp Val Pro Val Pro Ser Phe Ser Asp Ser Asp
55                  60                  65                  70
cct gca gca att att cat gac ttt gaa aag gga atg act gct tac ctg   345
Pro Ala Ala Ile Ile His Asp Phe Glu Lys Gly Met Thr Ala Tyr Leu
                75                  80                  85
gac ttg ttg ctg ggg atc tgc tat ctg atg ccc ctc aat act tct att   393
Asp Leu Leu Leu Gly Ile Cys Tyr Leu Met Pro Leu Asn Thr Ser Ile
                90                  95                  100
gtt atg cct cca aaa aat ctg gta gag ctc ttt ggc aaa ctg gcg agt   441
Val Met Pro Pro Lys Asn Leu Val Glu Leu Phe Gly Lys Leu Ala Ser
                105                 110                 115
ggc aga tat ctg cct caa act tat gtg gtt cga gaa gac cta gtt gct   489
Gly Arg Tyr Leu Pro Gln Thr Tyr Val Val Arg Glu Asp Leu Val Ala
                120                 125                 130
```

```
gtg gag gaa att cgt gat gtt agt aac ctt ggc atc ttt att tac caa      537
Val Glu Glu Ile Arg Asp Val Ser Asn Leu Gly Ile Phe Ile Tyr Gln
135                 140                 145                 150

ctt tgc aat aac aga aag tcc ttc cgc ctt cgt cgc aga gac ctc ttg      585
Leu Cys Asn Asn Arg Lys Ser Phe Arg Leu Arg Arg Arg Asp Leu Leu
                155                 160                 165

ctg ggt ttc aac aaa cgt gcc att gat aaa tgc tgg aag att aga cac      633
Leu Gly Phe Asn Lys Arg Ala Ile Asp Lys Cys Trp Lys Ile Arg His
            170                 175                 180

ttc ccc aac gaa ttt att gtt gag acc aag atc tgt caa gag            675
Phe Pro Asn Glu Phe Ile Val Glu Thr Lys Ile Cys Gln Glu
        185                 190                 195

taagaggcaa cagatagagt gtccttggta ataagaagtc agagatttac aatatgactt     735
taacattaag gtttatggga tactcaagat atttactcat gcatttactc tattgcttat     795
gctttaaaaa aaggaaaaaa aaaaaactac taaccactgc aagctcttgt caaattttag     855
tttaattggc attgcttgtt ttttgaaact gaaattacat gagtttcatt ttttctttgc     915
atttataggg tttagaatgc agc atgaatatat cacctaacat cctgacaata            975
aattccatcc gttgtttttt ttgtttgttt gttttttctt ttcctttaag taagctcttt    1035
attcatctta tggtggagca attttaaaat ttgaaatatt ttaaattgtt tttgaacttt    1095
ttgtgtaaaa tatatcagat ctcaacattg ttggtttctt ttgttttttca ttttgtacaa    1155
ctttcttgaa tttagaaatt acatctttgc agttctgtta ggtgctctgt aattaacctg    1215
acttatatgt gaacaatttt catgagacag tcattttaa ctaatgcagt gattctttct    1275
cactactatc tgtattgtgg aatgcacaaa attgtgtagg tgctgaatgc tgtaaggagt    1335
ttaggttgta tgaattctac aaccctataa taaattttac tctatacaaa aaaaaaa      1392
```

```
<210> 131
<211> 999
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 62..385

<221> polyA_signal
<222> 974..979

<221> polyA_site
<222> 987..999

<400> 131
cctgaatgac ttgaatgttt ccccgcctga gctaacagtc catgtgggtg attcagctct      60
g atg gga tgt gtt ttc cag agc aca gaa gac aaa tgt ata ttc aag ata     109
  Met Gly Cys Val Phe Gln Ser Thr Glu Asp Lys Cys Ile Phe Lys Ile
  1               5                   10                  15

gac tgg act ctg tca cca gga gag cac gcc aag gac gaa tat gtg cta       157
Asp Trp Thr Leu Ser Pro Gly Glu His Ala Lys Asp Glu Tyr Val Leu
        20                  25                  30

tac tat tac tcc aat ctc agt gtg cct att ggg cgc ttc cag aac cgc       205
Tyr Tyr Tyr Ser Asn Leu Ser Val Pro Ile Gly Arg Phe Gln Asn Arg
            35                  40                  45

gta cac ttg atg ggg gac atc tta tgc aat gat ggc tct ctc ctg ctc       253
Val His Leu Met Gly Asp Ile Leu Cys Asn Asp Gly Ser Leu Leu Leu
        50                  55                  60

caa gat gtg caa gag gct gac cag gga acc tat atc tgt gaa atc cgc       301
Gln Asp Val Gln Glu Ala Asp Gln Gly Thr Tyr Ile Cys Glu Ile Arg
65                  70                  75                  80

ctc aaa ggg gag agc cag gtg ttc aag aag gcg gtg gta ctg cat gtg       349
Leu Lys Gly Glu Ser Gln Val Phe Lys Lys Ala Val Val Leu His Val
                85                  90                  95
```

```
ctt cca gag gag ccc aaa ggt acg caa atg ctt act taaagagggg        395
Leu Pro Glu Glu Pro Lys Gly Thr Gln Met Leu Thr
            100                 105
ccaaggggca agagctttca tgtgcaagag gcaaggaaac tgattatctt gagtaaatgc   455
cagcctttgg gctaagtact taccacagag tgaatcttca aaaaatgatc ataattattt   515
cagtcaataa aaatagagtt attttattaa ataaaatatt gataattatt gtattattac   575
tttaaacaca cttccccctc acaaaagccc tgtgaaggat gttttgttca catatatgtc   635
caaatatgtt ttggacacat atttattaaa tggaataaat agtacttgaa ccctggcacc   695
tctgacaaca aagtccatgt tcttttttact atgccctaat acctttcatc agttatccac   755
attgatgcta catctgtatt ttataggtac cctatgttag gtgttctggg ggatagaaaa   815
gaaataagca ggccaggctc agtggctcat gcctgtaatc ctagcatttt gggaggctga   875
ggcagcagaa ctgcctgagc cccagggttc aagactgcag tgagctatga tggcaccact   935
gcattctagc ctgggtgaca gagcaagact ctgtctaaaa taaaaaaga gaaaaaaaa    995
aaaa                                                              999
```

```
<210> 132
<211> 725
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 422..550

<221> sig_peptide
<222> 422..475
<223> Von Heijne matrix
      score 4.5
      seq LRWLMPVIPALWG/AE

<221> polyA_site
<222> 714..725

<400> 132
tctgcgaggg tgggagagaa aattaggggg agaaaggaca gagagagcaa ctaccatcca   60
tagccagata ggtgagtaaa tatatttgca gtaacctatt tgctattcct tgctgcaact   120
gtgtttaatg ttccttccag aatcagagag agtattgcca tccaagaaat cgtttttaga   180
tatgacattt gagctatcat cttgagacca atacctaaaa caattttcagt ttaagaaatg   240
tctaggtatg gtgaaaacac agtttaaaaac cagcaaaaca gaatttattg ccctcagcga   300
atacccacaa tgtacatata ccttgtattt ctgaaagcaa agcaagcatg ccaagtagtt   360
tttatttacc tgtacctata atacagcaag gtgaaacagg atatattttt gaagtttaaa   420
a atg tct tca ggc cgg ctg cgg tgg ctc atg cct gta atc cca gca ctt  469
  Met Ser Ser Gly Arg Leu Arg Trp Leu Met Pro Val Ile Pro Ala Leu
          -15                 -10                 -5
tgg gga gcc gag aag ggt gaa tca cct gag gtc agc agt ttt gag acc   517
Trp Gly Ala Glu Lys Gly Glu Ser Pro Glu Val Ser Ser Phe Glu Thr
        1               5                   10
agg ctg gcc aac atg gcg aaa ccc tgt ctc tac tgaaaataca aaaattagct   570
Arg Leu Ala Asn Met Ala Lys Pro Cys Leu Tyr
15                  20                  25
gggtgtggtg gcgggcgcct gtagtcccag ctacttggga gactgaggca ggagaattgc   630
ttgaacacgg aaggcggaag ttgcagtaag ctgagatcgt gccaccgcac accagcttgg   690
gcaacagagt gagactccct ctcaaaaaaa aaaaa                             725
```

```
<210> 133
<211> 400
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 124..231

<221> polyA_site
<222> 387..400

<400> 133
ctcgcctctc ctggcttctg gtatgcacca gcaattcctg gcgttccttg gctcctagaa      60
gcatcactcc tatcacatgg tcatcttcac cctgtgtgtc ttcacactac cctttctctg     120
tgc atg tct gcc cga atc cct ttt tat aag gac acc agt cag att aga       168
    Met Ser Ala Arg Ile Pro Phe Tyr Lys Asp Thr Ser Gln Ile Arg
    1               5                   10                  15
tta ggg tct acc ata ata cct cat ttt aac tta atc acc ttt gta aag       216
Leu Gly Ser Thr Ile Ile Pro His Phe Asn Leu Ile Thr Phe Val Lys
                20                  25                  30
acc ttt ttc caa ata tagtcactct ctgaggtact gatggttagg atctcaacat       271
Thr Phe Phe Gln Ile
            35
accttttttg ggaggacaca attgaaccca taacagggtg tttgcaagga agagttaaaa     331
tttgaaagaa aggtggtatt tgcttagata gatagggcac agctttctag gtgacaaaaa     391
aaaaaaaaa                                                            400
```

```
<210> 134
<211> 1053
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 131..1051

<221> sig_peptide
<222> 131..169
<223> Von Heijne matrix
      score 4.2
      seq MLAVSLTVPLLGA/MM

<221> polyA_signal
<222> 1019..1024

<400> 134
gagcgaggcg gacgggctgc gacagcgccg gcccctgcgg ccgcaggtcg tcacagacga      60
tgatggccag gccccggagg ctaaggacgg cagctccttt agcggcagag ttttccgagt     120
gaccttcttg atg ctg gct gtt tct ctc acc gtt ccc ctg ctt gga gcc        169
                Met Leu Ala Val Ser Leu Thr Val Pro Leu Leu Gly Ala
                    -10                 -5
atg atg ctg ctg gaa tct cct ata gat cca cag cct ctc agc ttc aaa       217
Met Met Leu Leu Glu Ser Pro Ile Asp Pro Gln Pro Leu Ser Phe Lys
1               5                   10                  15
gaa ccc ccg ctc ttg ctt ggt gtt ctg cat cca aat acg aag ctg cga       265
Glu Pro Pro Leu Leu Leu Gly Val Leu His Pro Asn Thr Lys Leu Arg
            20                  25                  30
cag gca gaa agg ctg ttt gaa aat caa ctt gtt gga ccg gag tcc ata       313
Gln Ala Glu Arg Leu Phe Glu Asn Gln Leu Val Gly Pro Glu Ser Ile
        35                  40                  45
gca cat att ggg gat gtg atg ttt act ggg aca gca gat ggc cgg gtc       361
Ala His Ile Gly Asp Val Met Phe Thr Gly Thr Ala Asp Gly Arg Val
    50                  55                  60
```

```
gta aaa ctt gaa aat ggt gaa ata gag acc att gcc cgg ttt ggt tcg      409
Val Lys Leu Glu Asn Gly Glu Ile Glu Thr Ile Ala Arg Phe Gly Ser
65                  70                  75                  80
ggc cct tgc aaa acc cga gat gat gag cct gtg tgt ggg aga ccc ctg      457
Gly Pro Cys Lys Thr Arg Asp Asp Glu Pro Val Cys Gly Arg Pro Leu
                85                  90                  95
ggt atc cgt gca ggg ccc aat ggg act ctc ttt gtg gcc gat gca tgc      505
Gly Ile Arg Ala Gly Pro Asn Gly Thr Leu Phe Val Ala Asp Ala Cys
                100                 105                 110
aag gga cta ttt gaa gta aat ccc tgg aaa cgt gaa gtg aaa ctg ctg      553
Lys Gly Leu Phe Glu Val Asn Pro Trp Lys Arg Glu Val Lys Leu Leu
        115                 120                 125
ctg tcc tcc gag aca ccc att gag ggg aag aac atg tcc ttt gtg aat      601
Leu Ser Ser Glu Thr Pro Ile Glu Gly Lys Asn Met Ser Phe Val Asn
        130                 135                 140
gat ctt aca gtc tct cag gat ggg agg aag att tat ttc acc gat tct      649
Asp Leu Thr Val Ser Gln Asp Gly Arg Lys Ile Tyr Phe Thr Asp Ser
145             .   150                 155                 160
agc agc aaa tgg caa aga cga gac tac ctg ctt ctg gtg atg gag ggc      697
Ser Ser Lys Trp Gln Arg Arg Asp Tyr Leu Leu Leu Val Met Glu Gly
                165                 170                 175
aca gat gac ggg cgc ctg ctg gag tat gat act gtg acc agg gaa gta      745
Thr Asp Asp Gly Arg Leu Leu Glu Tyr Asp Thr Val Thr Arg Glu Val
                180                 185                 190
aaa gtt tta ttg gac cag ctg cgg ttc ccg aat gga gtc cag ctg tct      793
Lys Val Leu Leu Asp Gln Leu Arg Phe Pro Asn Gly Val Gln Leu Ser
        195                 200                 205
cct gca gaa gac ttt gtc ctg gtg gca gaa aca acc atg gcc agg ata      841
Pro Ala Glu Asp Phe Val Leu Val Ala Glu Thr Thr Met Ala Arg Ile
        210                 215                 220
cga aga gtc tac gtt tct ggc ctg atg aag ggc ggg gct gat ctg ttt      889
Arg Arg Val Tyr Val Ser Gly Leu Met Lys Gly Gly Ala Asp Leu Phe
225             .   230                 235                 240
gtg gag aac atg cct gga ttt cca gac aac atc cgg ccc agc agc tct      937
Val Glu Asn Met Pro Gly Phe Pro Asp Asn Ile Arg Pro Ser Ser Ser
                245                 250                 255
ggg ggg tac tgg gtg ggc atg tcg acc atc cgc cct aac cct ggg ttt      985
Gly Gly Tyr Trp Val Gly Met Ser Thr Ile Arg Pro Asn Pro Gly Phe
                260                 265                 270
tcc atg ctg gat ttc tta tct gag aga ccc tgg att aaa agg atg att     1033
Ser Met Leu Asp Phe Leu Ser Glu Arg Pro Trp Ile Lys Arg Met Ile
        275                 280                 285
ttt aag gca aaa aaa aaa aa                                          1053
Phe Lys Ala Lys Lys Lys
        290
```

```
<210> 135
<211> 1128
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 86..403

<221> sig_peptide
<222> 86..181
<223> Von Heijne matrix
      score 8.8   ·
      seq VPMLLLIVGGSFG/LR
```

<221> polyA_signal
<222> 1097..1102

<221> polyA_site
<222> 1117..1128

<400> 135
```
cgtcttggtg agagcgtgag ctgctgagat ttgggagtct gcgctaggcc cgcttggagt        60
tctgagccga tggaagagtt cactc atg ttt gca ccc gcg gtg atg cgt gct       112
                                 Met Phe Ala Pro Ala Val Met Arg Ala
                                     -30              -25
ttt cgc aag aac aag act ctc ggc tat gga gtc ccc atg ttg ttg ctg       160
Phe Arg Lys Asn Lys Thr Leu Gly Tyr Gly Val Pro Met Leu Leu Leu
        -20              -15              -10
att gtt gga ggt tct ttt ggt ctt cgt gag ttt tct caa atc cga tat       208
Ile Val Gly Gly Ser Phe Gly Leu Arg Glu Phe Ser Gln Ile Arg Tyr
        -5                   1                5
gat gct gtg aag agt aaa atg gat cct gag ctt gaa aaa aaa ctg aaa       256
Asp Ala Val Lys Ser Lys Met Asp Pro Glu Leu Glu Lys Lys Leu Lys
10                   15                   20                   25
gag aat aaa ata tct tta gag tcg gaa tat gag aaa atc aaa gac tcc       304
Glu Asn Lys Ile Ser Leu Glu Ser Glu Tyr Glu Lys Ile Lys Asp Ser
                     30                   35                   40
aag ttt gat gac tgg aag aat att cga gga ccc agg cct tgg gaa gat       352
Lys Phe Asp Asp Trp Lys Asn Ile Arg Gly Pro Arg Pro Trp Glu Asp
             45                   50                   55
cct gac ctc ctc caa gga aga aat cca gaa agc ctt aag act aag aca       400
Pro Asp Leu Leu Gln Gly Arg Asn Pro Glu Ser Leu Lys Thr Lys Thr
             60                   65                   70
act tgactctgct gattcttttt tccnnntttt tttttttta aataaaaata            453
Thr
ctattaactg gacttcctaa tatatacttc tatcaagtgg aaaggaaatt ccaggcccat       513
ggaaacttgg atatgggtaa tttgatgaca aataatcttc actaaaggtc atgtacaggt       573
tttttatactt cccagctatt ccatctgtgg atgaaagtaa caatgttggc cacgtatatt      633
ttacacctcg aaataaaaaa tgtgaatact gctccaaaaa aaaaaaccag taccgtgtag       693
tctctctcgt ggcttggatt tacactgggc aacgtggttg gaatgtatct ggctcagaac       753
tatgatatac caaacctggc taaaaaactt gaagaaatta aaaaggactt ggatgccaag       813
aagaaacccc ctagtgcatg agactgcctc cagcactgcc ttcaggatat accgattcta       873
ctgctcttga gggcctcgtt tactatctga accaaaagct tttgtttttcg tctccagcct      933
cagcacttct cttctttgct agaccctgtg tttttttgctt taaagcaagc aaaatggggc      993
cccaatttga gaactacccg acgtttccaa catactcacc tcttcccata atccctttcc      1053
aactgcatgg gaggttctaa gactggaatt atggtgctag attagtaaac atgacttta      1113
acgaaaaaaa aaaaa                                                        1128
```

<210> 136
<211> 254
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 37..162

<221> sig_peptide
<222> 37..93
<223> Von Heijne matrix
      score 9.5
      seq LMCLSLCTAFALS/KP

```
<221> polyA_signal
<222> 224..229

<221> polyA_site
<222> 243..254

<400> 136
tgtgctgtgg gggctacgag gaaagatcta attatc atg gac ctg cga cag ttt        54
                                        Met Asp Leu Arg Gln Phe
                                                            -15
ctt atg tgc ctg tcc ctg tgc aca gcc ttt gcc ttg agc aaa ccc aca       102
Leu Met Cys Leu Ser Leu Cys Thr Ala Phe Ala Leu Ser Lys Pro Thr
        -10             -5                   1
gaa aag aag gac cgt gta cat cat gag cct cag ctc agt gac aag gtt       150
Glu Lys Lys Asp Arg Val His His Glu Pro Gln Leu Ser Asp Lys Val
 5                   10                  15
cac aat gat att tgatagaacc aattgttgta cataaaacag atctgcgcat           202
His Asn Asp Ile
20
atatatatat gtataaaaaa taataaaata atggaagatg aaaaaaaaaa aa             254
```

```
<210> 137
<211> 886
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 31..381

<221> sig_peptide
<222> 31..90
<223> Von Heijne matrix
      score 5.4
      seq AFVIACVLSLIST/IY

<221> polyA_site
<222> 875..886

<400> 137
ggaggatggg cgagcagtct gaatggcaga atg gat aac cgt ttt gct aca gca      54
                                    Met Asp Asn Arg Phe Ala Thr Ala
                                    -20             -15
ttt gta att gct tgt gtg ctt agc ctc att tcc acc atc tac atg gca      102
Phe Val Ile Ala Cys Val Leu Ser Leu Ile Ser Thr Ile Tyr Met Ala
        -10             -5                   1
gcc tcc att ggc aca gac ttc tgg tat gaa tat cga agt cca gtt caa      150
Ala Ser Ile Gly Thr Asp Phe Trp Tyr Glu Tyr Arg Ser Pro Val Gln
5                   10                  15                  20
gaa aat tcc agt gat ttg aat aaa agc atc tgg gat gaa ttc att agt      198
Glu Asn Ser Ser Asp Leu Asn Lys Ser Ile Trp Asp Glu Phe Ile Ser
                25                  30                  35
gat gag gca gat gaa aag act tat aat gat gca ctt ttt cga tac aat      246
Asp Glu Ala Asp Glu Lys Thr Tyr Asn Asp Ala Leu Phe Arg Tyr Asn
                40                  45                  50
ggc aca gtg gga ttg tgg gga cgg tgt atc acc ata ccc aaa aac atg      294
Gly Thr Val Gly Leu Trp Gly Arg Cys Ile Thr Ile Pro Lys Asn Met
        55                  60                  65
cat tgg tat agc cca cca gaa agg aca ggt att tct ctt att tta act      342
His Trp Tyr Ser Pro Pro Glu Arg Thr Gly Ile Ser Leu Ile Leu Thr
```

```
           70                    75                    80
tct gtc ttc ttc acc tgg tta ata ata gac aaa acg acg taatgattgc       391
Ser Val Phe Phe Thr Trp Leu Ile Ile Asp Lys Thr Thr
85                    90                    95
ccaattacat gtaagcaggt ttgttggttc tctctctcct taaagaaata aatcgtgtat    451
cttctctttc tactgccttc tctccccaac ttctttgcat taccatggta ctcatcaata    511
ttggttggat gaggaacttt tcttatcttg ggaaagcctt aatggctttt ttttttctta    571
tttactcact cattaaaata cttttcatta ctctaacaca tgttataaag aaatagttgg    631
aaaagtgcat cgaaagactt ttaaaaatat ttggtaacta gtaaaggac taccatcgaa     691
aatcaactca aaaaattgtc cttttatggg ttagctgtat tataatacat atctatcatt    751
tgcccctgtg tcttagagga tataatttga ccagctctac atttaatctg tgtaattatg    811
agactgtttt acaacaatct tgatgcagag ttggtaggtt aagaaattg tattacagaa     871
gttaaaaaaa aaaaa                                                      886
```

```
<210> 138
<211> 1244
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 46..579

<221> sig_peptide
<222> 46..156
<223> Von Heijne matrix
      score 3.5
      seq LVFNFLLILTILT/IW

<400> 138
cccttatcca ggttnttatc tanggaatcc cnnnaagact ggggа atg gag aga cag    57
                                                    Met Glu Arg Gln
                                                             -35
tca agg gtt atg tca gaa aag gat gag tat cag ttt caa cat can nna      105
Ser Arg Val Met Ser Glu Lys Asp Glu Tyr Gln Phe Gln His Xaa Xaa
         -30                 -25                 -20
gcg gng gan ctg ctt gtc ttc aat ttt ttg ctc atc ctt acc att ttg     153
Ala Xaa Xaa Leu Leu Val Phe Asn Phe Leu Leu Ile Leu Thr Ile Leu
     -15                 -10                  -5
aca atc tgg tta ttt aaa aat cat cga ttc cgc ttc ttg cat gaa act     201
Thr Ile Trp Leu Phe Lys Asn His Arg Phe Arg Phe Leu His Glu Thr
 1               5                  10                  15
gga gga gca atg gtg tat ggc ctt ata atg gga cta att tca cga tat     249
Gly Gly Ala Met Val Tyr Gly Leu Ile Met Gly Leu Ile Ser Arg Tyr
             20                  25                  30
gct aca gca cca act gat att gaa agt gga act gtc tgt gac tgt gta     297
Ala Thr Ala Pro Thr Asp Ile Glu Ser Gly Thr Val Cys Asp Cys Val
         35                  40                  45
aaa cta act ttc agt cca cca act ctg ctg gtt aat gtc act gac caa     345
Lys Leu Thr Phe Ser Pro Pro Thr Leu Leu Val Asn Val Thr Asp Gln
         50                  55                  60
gtt tat gaa tat aaa tac aaa aga gaa ata agt cag cac aac atc aat     393
Val Tyr Glu Tyr Lys Tyr Lys Arg Glu Ile Ser Gln His Asn Ile Asn
         65                  70                  75
cct cat caa gga aat gct ata ctt gaa aag atg aca ttt gat cca gaa     441
Pro His Gln Gly Asn Ala Ile Leu Glu Lys Met Thr Phe Asp Pro Glu
80                  85                  90                  95
atc ttc ttc aat gtt tta ctg cca cca att ata ttt cat gca gga tat     489
Ile Phe Phe Asn Val Leu Leu Pro Pro Ile Ile Phe His Ala Gly Tyr
             100                 105                 110
```

```
agt cta aag aag aga cac ttt ttt caa aac tta gga tct att tta acg    537
Ser Leu Lys Lys Arg His Phe Phe Gln Asn Leu Gly Ser Ile Leu Thr
            115              120              125
tat gcc ttc ttg gga act gcc atc tcc tgc atc gtc ata ggg            579
Tyr Ala Phe Leu Gly Thr Ala Ile Ser Cys Ile Val Ile Gly
            130              135              140
taagtgacat tcggagctca agttgcaggt ggctgtgggg tctgtgatct gtgtgaggga   639
tctaacactt ccaggattct tgctggctgg gaaaattgtc tttttttag tatatcacat    699
atttgtatgt ttttctgac ttaattccac ggcttctgac aaatacaagg cttcaaatca    759
aagcaaacta gaggattgct ggactttctc tgtgagttct ggacttctga cttagggaat   819
gtggatcact tgccttgagt tatgtgaagc gcattgcatt cttcttttag tttgagtaat    879
gccgatatgg tcactgcatt ctttttgtc ttgtattgag agaccttacc tgtatttggc    939
aggagtgcaa aagtaactat atgccaagag tttctttct aaaggaaagt ttacaagaca    999
gcagtctgaa acagatatgn tccaaatatn naacagagtt gcttaataca gggatagctt   1059
ttcagttaat accctgtaga atgcagactc ttntttcat tgtattttct tgattatgct    1119
actgagccct aagtcacacg ttatatactc tggcttgcag ctcatcataa agtaaaatgt   1179
ggtaccaaat ggtgaaggca atccagcctn tgataatccc gtccaataca ttaaagntcc   1239
actgc                                                              1244
```


```
<210> 139
<211> 471
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 92..469

<221> sig_peptide
<222> 92..172
<223> Von Heijne matrix
      score 7.9
      seq VVVLALGFLGCYG/AK

<221> polyA_signal
<222> 454..459

<221> polyA_site
<222> 458..471

<400> 139
gcaagtgcag aagtcggtga cggtgggcat ctgggtgtca atcgatgggg catcctttct    60
gaagatcttc gggccactgt cgtccagtgc c atg cag ttt gtc aac gtg ggc     112
                                  Met Gln Phe Val Asn Val Gly
                                                         -25
tac ttc ctc atc gca gcc ggc gtt gtg gtc ctt gct ctt ggt ttc ctg   160
Tyr Phe Leu Ile Ala Ala Gly Val Val Val Leu Ala Leu Gly Phe Leu
-20              -15              -10                          -5
ggc tgc tat ggt gct aag act gag agc atg tgt gcc ctc gtg acg ttc   208
Gly Cys Tyr Gly Ala Lys Thr Glu Ser Met Cys Ala Leu Val Thr Phe
                  1                5                    10
ttc ttc atc ctc ctc ctc atc ttc att gct gag gtt gca gct gct gtg   256
Phe Phe Ile Leu Leu Leu Ile Phe Ile Ala Glu Val Ala Ala Ala Val
              15               20                   25
gtc gcc ctg gtg tac acc aca atg gct gag cac ttc ctg acg ttg ctg   304
Val Ala Leu Val Tyr Thr Thr Met Ala Glu His Phe Leu Thr Leu Leu
          30               35                   40
gta gtg cct gcc atc aag aaa gat tat ggt tcc cag gaa gac ttc act   352
Val Val Pro Ala Ile Lys Lys Asp Tyr Gly Ser Gln Glu Asp Phe Thr
45               50               55                   60
```

```
caa gtg tgg aac acc acc atg aaa ggg ctc aag tgc cgt ggc ttc acc    400
Gln Val Trp Asn Thr Thr Met Lys Gly Leu Lys Cys Arg Gly Phe Thr
                 65              70              75
aac tat acg gat ttt gag gac tca ccc tac ttc aaa atg cat aaa cct    448
Asn Tyr Thr Asp Phe Glu Asp Ser Pro Tyr Phe Lys Met His Lys Pro
             80              85              90
gtt aca atg aaa aaa aaa aaa aa                                      471
Val Thr Met Lys Lys Lys Lys
         95
```

```
<210> 140
<211> 849
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 154..675

<221> sig_peptide
<222> 154..498
<223> Von Heijne matrix
      score 4.8
      seq PLRLLNLLILIEG/GV

<221> polyA_signal
<222> 819..824

<221> polyA_site
<222> 838..849

<400> 140
```

```
ccctatctc cagacctcat tcgcaatgaa gtagaatgtc tgaaagcaga tttcaaccac     60
agaatcaagg aggttctctt caactccctc ttcagtgcct actatgttgc atttctcccc    120
ctgtgttttg tgaagagtac ccagtactat gac atg cgc tgg tca tgt gag cac    174
                                     Met Arg Trp Ser Cys Glu His
                                    -115              -110
ctc gtt atg gtg tgg atc aat gct ttt gtc atg ctc acc acg caa ctg    222
Leu Val Met Val Trp Ile Asn Ala Phe Val Met Leu Thr Thr Gln Leu
        -105             -100             -95
ttg cca tcc aaa tac tgt gat ttg cta cat aaa tca gct gct cac ctg    270
Leu Pro Ser Lys Tyr Cys Asp Leu Leu His Lys Ser Ala Ala His Leu
        -90             -85             -80
ggc aag tgg cag aag ttg gaa cat ggg tcc tac agc aat gct cca cag    318
Gly Lys Trp Gln Lys Leu Glu His Gly Ser Tyr Ser Asn Ala Pro Gln
    -75             -70             -65
cac att tgg tca gaa aat aca ata tgg cct caa ggg gtg ctg gtg cgg    366
His Ile Trp Ser Glu Asn Thr Ile Trp Pro Gln Gly Val Leu Val Arg
-60             -55             -50             -45
cac agc aga tgt tta tat aga gcc atg ggg cct tac aac gtg gca gtg    414
His Ser Arg Cys Leu Tyr Arg Ala Met Gly Pro Tyr Asn Val Ala Val
            -40             -35             -30
cct tca gat gta tct cat gcc cgc ttt tat ttc tta ttt cat cga cca    462
Pro Ser Asp Val Ser His Ala Arg Phe Tyr Phe Leu Phe His Arg Pro
        -25             -20             -15
tta agg ctg tta aat ctg ctc atc ctt att gag ggc ggt gtc gtc ttc    510
Leu Arg Leu Leu Asn Leu Leu Ile Leu Ile Glu Gly Gly Val Val Phe
        -10             -5               1
tat cag ctc tat tcc ttg ctg cgg tcg gag aag tgg aac cac aca ctt    558
Tyr Gln Leu Tyr Ser Leu Leu Arg Ser Glu Lys Trp Asn His Thr Leu
```

```
         5                    10                   15                   20
tcc atg gct ctc atc ctc ttc tgc aac tac tat gtt tta ttt aaa ctt      606
Ser Met Ala Leu Ile Leu Phe Cys Asn Tyr Tyr Val Leu Phe Lys Leu
                 25                   30                   35
ctc cgg gac aga ata gta tta ggc agg gca tac tcc tac cca ctc aac      654
Leu Arg Asp Arg Ile Val Leu Gly Arg Ala Tyr Ser Tyr Pro Leu Asn
                 40                   45                   50
agt tat gaa ctc aag gca aac taagctgcct ctcaacaatg agggagaact        705
Ser Tyr Glu Leu Lys Ala Asn
                 55
cagataaaaa tattttcata cgttctattt tttcttgtg  atttttataa atatttaaga   765
tgttttatat tttgtatact attatgtttt gaaagtcggg aagagtaagg gatattaaat   825
gtatccgtaa acaaaaaaaa aaaa                                          849
```

<210> 141
<211> 155
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -31..-1

<400> 141
```
Met Phe Thr Ser Thr Gly Ser Ser Gly Leu Tyr Lys Ala Pro Leu Ser
    -30                 -25                 -20
Lys Ser Leu Leu Leu Val Pro Ser Ala Leu Ser Leu Leu Leu Ala Leu
-15                 -10                 -5                       1
Leu Leu Pro His Cys Gln Lys Pro Phe Val Tyr Asp Leu His Ala Val
             5                   10                  15
Lys Asn Asp Phe Gln Ile Trp Arg Leu Ile Cys Gly Arg Ile Ile Cys
         20                  25                  30
Leu Asp Leu Lys Asp Thr Phe Cys Ser Ser Leu Leu Ile Tyr Asn Phe
     35                  40                  45
Arg Ile Phe Glu Arg Arg Tyr Gly Ser Arg Lys Phe Ala Ser Phe Leu
50                  55                  60                      65
Leu Gly Thr Trp Val Leu Ser Ala Leu Phe Asp Phe Leu Leu Ile Glu
                 70                  75                  80
Ala Met Gln Tyr Phe Phe Gly Ile Thr Ala Ala Ser Asn Leu Pro Ser
             85                  90                  95
Gly Leu Ile Phe Cys Cys Ala Phe Cys Ser Glu Thr Lys Leu Phe Leu
         100                 105                 110
Ser Arg Gln Ala Met Ala Glu Asn Phe Ser Ile
    115                 120
```

<210> 142
<211> 55
<212> PRT
<213> Homo sapiens

<400> 142
```
Met Ala Asp Phe Tyr Lys Glu Phe Leu Ser Lys Asn Phe Gln Lys Arg
1               5                   10                  15
Met Tyr Tyr Asn Arg Asp Trp Tyr Lys Arg Asn Phe Ala Ile Thr Phe
             20                  25                  30
Phe Met Gly Lys Val Ala Leu Glu Arg Ile Trp Asn Lys Leu Lys Gln
         35                  40                  45
Lys Gln Lys Lys Arg Ser Asn
```

50   55

<210> 143
<211> 67
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -20..-1

<400> 143
Met Ser Arg Asn Leu Arg Thr Ala Leu Ile Phe Gly Gly Phe Ile Ser
-20   -15   -10   -5
Leu Ile Gly Ala Ala Phe Tyr Pro Ile Tyr Phe Arg Pro Leu Met Arg
           1   5   10
Leu Glu Glu Tyr Lys Lys Glu Gln Ala Ile Asn Arg Ala Gly Ile Val
       15   20   25
Gln Glu Asp Val Gln Pro Pro Gly Leu Lys Val Trp Ser Asp Pro Phe
   30   35   40
Gly Arg Lys
45

<210> 144
<211> 198
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 144
Met Pro Val Pro Ala Leu Cys Leu Leu Trp Ala Leu Ala Met Val Thr
   -20   -15   -10
Arg Pro Ala Ser Ala Ala Pro Met Gly Gly Pro Glu Leu Ala Gln His
-5   1   5   10
Glu Glu Leu Thr Leu Leu Phe His Gly Thr Leu Gln Leu Gly Gln Ala
           15   20   25
Leu Asn Gly Val Tyr Arg Thr Thr Glu Gly Trp Leu Thr Lys Ala Arg
       30   35   40
Asn Ser Leu Gly Leu Tyr Gly Arg Thr Ile Glu Leu Leu Gly Gln Glu
   45   50   55
Val Ser Arg Gly Arg Asp Ala Ala Gln Glu Leu Arg Ala Ser Leu Leu
60   65   70   75
Glu Thr Gln Met Glu Glu Asp Ile Leu Gln Leu Gln Ala Glu Ala Thr
               80   85   90
Ala Glu Val Leu Gly Glu Val Ala Gln Ala Gln Lys Val Leu Arg Asp
           95   100   105
Ser Val Gln Arg Leu Glu Val Gln Leu Arg Ser Ala Trp Leu Gly Pro
       110   115   120
Ala Tyr Arg Glu Phe Glu Val Leu Lys Ala His Ala Asp Lys Gln Ser
       125   130   135
His Ile Leu Trp Ala Leu Thr Gly His Val Gln Arg Gln Arg Arg Glu
140   145   150   155
Met Val Ala Gln Gln His Arg Leu Arg Gln Ile Gln Glu Arg Leu His
               160   165   170
Thr Ala Ala Leu Pro Ala

175

<210> 145
<211> 135
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -25..-1

<400> 145
Met Ser Leu Arg Asn Leu Trp Arg Asp Tyr Lys Val Leu Val Val Met
-25   -20   -15   -10
Val Pro Leu Val Gly Leu Ile His Leu Gly Trp Tyr Arg Ile Lys Ser
              -5   1   5
Ser Pro Val Phe Gln Ile Pro Lys Asn Asp Asp Ile Pro Glu Gln Asp
         10   15   20
Ser Leu Gly Leu Ser Asn Leu Gln Lys Ser Gln Ile Gln Gly Lys Xaa
    25   30   35
Ala Gly Leu Gln Ser Ser Gly Lys Glu Ala Ala Leu Asn Leu Ser Phe
40   45   50   55
Ile Ser Lys Glu Glu Met Lys Asn Thr Ser Trp Ile Arg Lys Asn Trp
         60   65   70
Leu Leu Val Ala Gly Ile Ser Phe Ile Gly Asp His Leu Gly Thr Tyr
         75   80   85
Phe Leu Gln Arg Ser Ala Lys Gln Ser Val Lys Phe Gln Ser Gln Ser
         90   95   100
Lys Gln Lys Ser Ile Glu Glu
    105   110

<210> 146
<211> 255
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -70..-1

<400> 146
Met Gln Gln Lys Glu Gln Gln Phe Arg Glu Trp Phe Leu Lys Glu Phe
-70   -65   -60   -55
Pro Gln Ile Arg Trp Lys Ile Gln Glu Ser Ile Glu Arg Leu Arg Val
         -50   -45   -40
Ile Ala Asn Glu Ile Glu Lys Val His Arg Gly Cys Val Ile Ala Asn
         -35   -30   -25
Val Val Ser Gly Ser Thr Gly Ile Leu Ser Val Ile Gly Val Met Leu
         -20   -15   -10
Ala Pro Phe Thr Ala Gly Leu Ser Leu Ser Ile Thr Ala Ala Gly Val
    -5   1   5   10
Gly Leu Gly Ile Ala Ser Ala Thr Ala Gly Ile Ala Ser Ser Ile Val
              15   20   25
Glu Asn Thr Tyr Thr Arg Ser Ala Glu Leu Thr Ala Ser Arg Leu Thr
         30   35   40
Ala Thr Ser Thr Asp Gln Leu Glu Ala Leu Arg Asp Ile Leu His Asp
    45   50   55
Ile Thr Pro Asn Val Leu Ser Phe Ala Leu Asp Phe Asp Glu Ala Thr

```
     60                        65                        70
Lys Met Ile Ala Asn Asp Val His Thr Leu Arg Arg Ser Lys Ala Thr
75                        80                        85                        90
Val Gly Arg Pro Leu Ile Ala Trp Arg Tyr Val Pro Ile Asn Val Val
                    95                        100                       105
Glu Thr Leu Arg Thr Arg Gly Ala Pro Thr Arg Ile Val Arg Lys Val
               110                       115                       120
Ala Arg Asn Leu Gly Lys Ala Thr Ser Gly Val Leu Val Val Leu Asp
          125                       130                       135
Val Val Asn Leu Val Gln Asp Ser Leu Asp Leu His Lys Gly Glu Lys
     140                       145                       150
Ser Glu Ser Ala Glu Leu Leu Arg Gln Trp Ala Gln Glu Leu Glu Glu
155                       160                       165                       170
Asn Leu Asn Glu Leu Thr His Ile His Gln Ser Leu Lys Ala Gly
               175                       180                       185
```

```
<210> 147
<211> 59
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -49..-1

<400> 147
Met Pro Gly Thr Glu Val Leu Glu Gly Ala Thr Asp Gly Leu Ala Ala
               -45                       -40                       -35
Ile Asn Leu Leu Lys Trp Ile Lys Thr Leu Gly Gly Ser Val Ile Ser
          -30                       -25                       -20
Met Ile Val Leu Leu Ile Cys Val Val Cys Leu Tyr Ile Val Cys Arg
          -15                       -10                       -5
Cys Gly Ser His Leu Trp Arg Glu Ser His His
     1                    5                         10
```

```
<210> 148
<211> 180
<212> PRT
<213> Homo sapiens

<400> 148
Met Cys Ile Ser Gly Leu Cys Gln Ile Val Gly Cys Asp His Gln Leu
1                    5                         10                        15
Gly Ser Thr Val Lys Glu Asp Asn Cys Gly Val Cys Asn Gly Asp Gly
               20                        25                        30
Ser Thr Cys Arg Leu Val Arg Gly Gln Tyr Lys Ser Gln Leu Ser Ala
          35                        40                        45
Thr Lys Ser Asp Asp Thr Val Val Ala Ile Pro Tyr Gly Ser Arg His
     50                        55                        60
Ile Arg Leu Val Leu Lys Gly Pro Asp His Leu Tyr Leu Glu Thr Lys
65                        70                        75                        80
Thr Leu Gln Gly Thr Lys Gly Glu Asn Ser Leu Ser Ser Thr Gly Thr
               85                        90                        95
Phe Leu Val Asp Asn Ser Ser Val Asp Phe Gln Lys Phe Pro Asp Lys
               100                       105                       110
Glu Ile Leu Arg Met Ala Gly Pro Leu Thr Ala Asp Phe Ile Val Lys
          115                       120                       125
Ile Arg Asn Ser Gly Ser Ala Asp Ser Thr Val Gln Phe Ile Phe Tyr
```

```
            130                   135                   140
Gln Pro Ile Ile His Arg Trp Arg Glu Thr Asp Phe Phe Pro Cys Ser
145                   150                   155                   160
Ala Thr Cys Gly Gly Gly Tyr Gln Leu Thr Ser Ala Glu Cys Tyr Asp
                165                   170                   175
Leu Arg Ser Asn
            180
```

```
<210> 149
<211> 162
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -23..-1

<400> 149
Met Gly Asp Lys Ile Trp Leu Pro Phe Pro Val Leu Leu Leu Ala Ala
            -20                   -15                   -10
Leu Pro Pro Val Leu Leu Pro Gly Ala Ala Gly Phe Thr Pro Ser Leu
        -5                    1                     5
Asp Ser Asp Phe Thr Phe Thr Leu Pro Ala Gly Gln Lys Glu Cys Phe
10                    15                    20                    25
Tyr Gln Pro Met Pro Leu Lys Ala Ser Leu Glu Ile Glu Tyr Gln Val
                30                    35                    40
Leu Asp Gly Ala Gly Leu Asp Ile Asp Phe His Leu Ala Ser Pro Glu
                45                    50                    55
Gly Lys Thr Leu Val Phe Glu Gln Arg Lys Ser Asp Gly Val His Thr
            60                    65                    70
Val Glu Thr Glu Val Gly Asp Tyr Met Phe Cys Phe Asp Asn Thr Phe
        75                    80                    85
Ser Thr Ile Ser Glu Lys Val Ile Phe Phe Glu Leu Ile Pro Asp Asn
90                    95                    100                   105
Met Gly Glu Gln Ala Gln Glu Gln Glu Asp Trp Lys Lys Tyr Ile Thr
                110                   115                   120
Gly Thr Asp Ile Leu Asp Met Lys Leu Glu Asp Ile Leu Val Ser Met
                125                   130                   135
Val Phe
```

```
<210> 150
<211> 120
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -23..-1

<400> 150
Met Gly Asp Lys Ile Trp Leu Pro Phe Pro Val Leu Leu Leu Ala Ala
            -20                   -15                   -10
Leu Pro Pro Val Leu Leu Pro Gly Ala Ala Gly Phe Thr Pro Ser Leu
        -5                    1                     5
Asp Ser Asp Phe Thr Phe Thr Leu Pro Ala Gly Gln Lys Glu Cys Phe
10                    15                    20                    25
Tyr Gln Pro Met Pro Leu Lys Ala Ser Leu Glu Ile Glu Tyr Gln Val
                30                    35                    40
```

Leu Asp Gly Ala Gly Leu Asp Ile Asp Phe His Leu Ala Ser Pro Glu
            45   50   55
Gly Lys Thr Leu Val Phe Glu Gln Arg Lys Ser Asp Gly Val His Thr
        60   65   70
Cys Ile Arg Ser Lys Asn Gly Pro Gly Thr Ala Val His Ala Tyr Asn
    75   80   85
Pro Ser Thr Phe Arg Gly Gln Val
90   95

<210> 151
<211> 7
<212> PRT
<213> Homo sapiens

<400> 151
Met Val Glu Met Thr Gly Val
1   5

<210> 152
<211> 199
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -42..-1

<400> 152
Met Asp Gly Gln Lys Lys Asn Trp Lys Asp Lys Val Val Asp Leu Leu
        -40   -35   -30
Tyr Trp Arg Asp Ile Lys Lys Thr Gly Val Val Phe Gly Ala Ser Leu
    -25   -20   -15
Phe Leu Leu Leu Ser Leu Thr Val Phe Ser Ile Val Ser Val Thr Ala
-10   -5   1   5
Tyr Ile Ala Leu Ala Leu Leu Ser Val Thr Ile Ser Phe Arg Ile Tyr
            10   15   20
Lys Gly Val Ile Gln Ala Ile Gln Lys Ser Asp Glu Gly His Pro Phe
        25   30   35
Arg Ala Tyr Leu Glu Ser Glu Val Ala Ile Ser Glu Glu Leu Val Gln
    40   45   50
Lys Tyr Ser Asn Ser Ala Leu Gly His Val Asn Cys Thr Ile Lys Glu
55   60   65   70
Leu Arg Arg Leu Phe Leu Val Asp Asp Leu Val Asp Ser Leu Lys Phe
                75   80   85
Ala Val Leu Met Trp Val Phe Thr Tyr Val Gly Ala Leu Phe Asn Gly
            90   95   100
Leu Thr Leu Leu Ile Leu Ala Leu Ile Ser Leu Phe Ser Val Pro Val
        105   110   115
Ile Tyr Glu Arg His Gln Ala Gln Ile Asp His Tyr Leu Val Leu Ala
    120   125   130
Asn Lys Asn Val Lys Asp Ala Met Ala Lys Ile Gln Ala Lys Ile Pro
135   140   145   150
Gly Leu Lys Arg Lys Ala Glu
                155

<210> 153

```
<211> 43
<212> PRT
<213> Homo sapiens

<400> 153
Met Pro Phe Arg Met Ser Gly Tyr Ile Pro Phe Gly Thr Pro Ile Val
1               5                   10                  15
Ser Val Thr Phe Lys Gly Phe Pro Phe Leu Lys Asn Tyr Phe Lys Cys
            20                  25                  30
Leu Thr Leu Cys Tyr Cys Ser Arg Val Phe Asp
            35                  40



<210> 154
<211> 50
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -37..-1

<400> 154
Met Glu Trp Ala Gly Lys Gln Arg Asp Phe Gln Val Arg Ala Ala Pro
        -35                 -30                 -25
Gly Trp Asp His Leu Ala Ser Phe Pro Gly Pro Ser Leu Arg Leu Phe
        -20         .       -15                 -10
Ser Gly Ser Gln Ala Ser Val Cys Ser Leu Cys Ser Gly Phe Gly Ala
-5                  .1              5                   10
Gln Glu



<210> 155
<211> 153
<212> PRT
<213> Homo sapiens

<400> 155
Thr Val Pro Leu Leu Leu Glu Pro Ala Asp His Ala Arg Gly Arg Ala
1               5                   10                  15
His Val His Leu Pro Glu Asn Val Arg Ser Gln Ser Pro Gly His Val
            20                  25                  30
Arg Arg Gly Arg Ser Gly Ala Gln Val Leu Pro Thr Gly Pro Asp Glu
            35                  40                  45
Lys Gln Val Glu Lys Ser Glu Val Asp Phe Ser Lys Ser His Ser Leu
        50                  55                  60
Val Arg Arg Phe Glu Asp Leu Lys Pro Lys Leu Ser Val Cys Lys Thr
65                  70                  75                  80
Gly Ser Gln Val Phe Arg Ser Glu Asn Trp Lys Val Trp Ala Glu Ser
                85                  90                  95
Ser Arg Gly Asp His Asp Asp Cys Leu Asp Leu Cys Ser Val Leu Cys
            100                 105                 110
Trp Gly Glu Leu Leu Arg Thr Ile Pro Glu Ile Pro Pro Lys Arg Gly
            115                 120                 125
Glu Leu Lys Thr Glu Leu Leu Gly Leu Lys Glu Arg Lys His Lys Pro
        130                 135                 140
Gln Val Ser Gln Gln Glu Glu Leu Lys
145                 150
```

```
<210> 156
<211> 67
<212> PRT
<213> Homo sapiens

<400> 156
Met Arg Gln Lys Arg Lys Gly Asp Leu Ser Pro Ala Lys Leu Met Met
1               5                   10                  15
Leu Thr Ile Gly Asp Val Ile Lys Gln Leu Ile Glu Ala His Glu Gln
            20                  25                  30
Gly Lys Asp Ile Asp Leu Asn Lys Val Arg Thr Lys Thr Ala Ala Lys
        35                  40                  45
Tyr Gly Leu Ser Ala Gln Pro Arg Leu Val Asp Ile Ile Ala Ala Val
    50                  55                  60
Pro Pro Glu
65
```

```
<210> 157
<211> 87
<212> PRT
<213> Homo sapiens

<400> 157
Met Asp Glu Leu Ser Glu Glu Asp Lys Leu Thr Val Ser Arg Ala Arg
1               5                   10                  15
Lys Ile Gln Arg Phe Leu Ser Gln Pro Phe Gln Val Ala Glu Val Phe
            20                  25                  30
Thr Gly His Met Gly Lys Leu Val Pro Leu Lys Glu Thr Ile Lys Gly
        35                  40                  45
Phe Gln Gln Ile Leu Ala Gly Glu Tyr Asp His Leu Pro Glu Gln Ala
    50                  55                  60
Phe Tyr Met Val Gly Pro Ile Glu Glu Ala Val Ala Lys Ala Asp Lys
65                  70                  75                  80
Leu Ala Glu Glu His Ser Ser
                85
```

```
<210> 158
<211> 250
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -85..-1

<400> 158
Met Ser Ala Glu Val Lys Val Thr Gly Gln Asn Gln Glu Gln Phe Leu
-85             -80                 -75                 -70
Leu Leu Ala Lys Ser Ala Lys Gly Ala Ala Leu Ala Thr Leu Ile His
            -65                 -60                 -55
Gln Val Leu Glu Ala Pro Gly Val Tyr Val Phe Gly Glu Leu Leu Asp
        -50                 -45                 -40
Met Pro Asn Val Arg Glu Leu Xaa Ala Arg Asn Leu Pro Pro Leu Thr
    -35                 -30                 -25
Glu Ala Gln Lys Asn Lys Leu Arg His Leu Ser Val Val Thr Leu Ala
    -20                 -15                 -10
Ala Lys Val Lys Cys Ile Pro Tyr Ala Val Leu Leu Glu Ala Leu Ala
```

```
-5                    1              5                   10
Leu Arg Asn Val Arg Gln Leu Glu Asp Leu Val Ile Glu Ala Val Tyr
                15                  20                  25
Ala Asp Val Leu Arg Gly Ser Leu Asp Gln Arg Asn Gln Arg Leu Glu
            30                  35                  40
Val Asp Tyr Ser Ile Gly Arg Asp Ile Gln Arg Gln Asp Leu Ser Ala
            45                  50                  55
Ile Ala Arg Thr Leu Gln Glu Trp Cys Val Gly Cys Glu Val Val Leu
60                  65                  70                  75
Ser Gly Ile Glu Glu Gln Val Ser Arg Ala Asn Gln His Lys Glu Gln
                80                  85                  90
Gln Leu Gly Leu Lys Gln Gln Ile Glu Ser Glu Val Ala Asn Leu Lys
            95                  100                 105
Lys Thr Ile Lys Val Thr Thr Ala Ala Ala Ala Ala Ala Thr Ser Gln
            110                 115                 120
Asp Pro Glu Gln His Leu Thr Glu Leu Arg Glu Pro Ala Pro Gly Thr
            125                 130                 135
Asn Gln Arg Gln Pro Ser Lys Lys Ala Ser Lys Gly Lys Gly Leu Arg
140                 145                 150                 155
Gly Ser Ala Lys Ile Trp Ser Lys Ser Asn
            160                 165
```

```
<210> 159
<211> 24
<212> PRT
<213> Homo sapiens

<400> 159
Met Pro Thr Asn Cys Ala Ala Ala Gly Cys Ala Thr Thr Tyr Asn Lys
1                5                   10                  15
His Ile Asn Ile Ser Phe His Arg
            20
```

```
<210> 160
<211> 228
<212> PRT
<213> Homo sapiens

<400> 160
Met Pro Thr Asn Cys Ala Ala Ala Gly Cys Ala Thr Thr Tyr Asn Lys
1                5                   10                  15
His Ile Asn Ile Ser Phe His Arg Phe Pro Leu Asp Pro Lys Arg Arg
            20                  25                  30
Lys Glu Trp Val Arg Leu Val Arg Arg Lys Asn Phe Val Pro Gly Lys
            35                  40                  45
His Thr Phe Leu Cys Ser Lys His Phe Glu Ala Ser Cys Phe Asp Leu
            50                  55                  60
Thr Gly Gln Thr Arg Arg Leu Lys Met Asp Ala Val Pro Thr Ile Phe
65                  70                  75                  80
Asp Phe Cys Thr His Ile Lys Ser Met Lys Leu Lys Ser Arg Asn Leu
                85                  90                  95
Leu Lys Lys Asn Asn Ser Cys Ser Pro Ala Gly Pro Ser Ser Leu Lys
            100                 105                 110
Ser Asn Ile Ser Ser Gln Gln Val Leu Leu Glu His Ser Tyr Ala Phe
            115                 120                 125
Arg Asn Pro Met Glu Ala Lys Lys Arg Ile Ile Lys Leu Glu Lys Glu
            130                 135                 140
Ile Ala Ser Leu Arg Arg Lys Met Lys Thr Cys Leu Gln Lys Glu Arg
```

```
          145                      150                      155                      160
Arg Ala Thr Arg Arg Trp Ile Lys Ala Met Cys Leu Val Lys Asn Leu
                     165                      170                      175
Glu Ala Asn Ser Val Leu Pro Lys Gly Thr Ser Glu His Met Leu Pro
              180                      185                      190
Thr Ala Leu Ser Ser Leu Pro Leu Glu Asp Phe Lys Ile Leu Glu Gln
              195                      200                      205
Asp Gln Gln Asp Lys Thr Leu Leu Ser Leu Asn Leu Lys Gln Thr Lys
         210                      215                      220
Ser Thr Phe Ile
225
```

<210> 161
<211> 86
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -20..-1

<400> 161
```
Met Asn Leu His Phe Pro Gln Trp Phe Val His Ser Ser Ala Leu Gly
-20                      -15                      -10                       -5
Leu Val Leu Ala Pro Pro Phe Ser Ser Pro Gly Thr Asp Pro Thr Phe
                  1                        5                       10
Pro Cys Ile Tyr Cys Arg Leu Leu Asn Met Ile Met Thr Arg Leu Ala
             15                       20                       25
Phe Ser Phe Ile Thr Cys Leu Cys Pro Asn Leu Lys Glu Val Cys Leu
         30                       35                       40
Ile Leu Pro Glu Lys Asn Cys Asn Ser Arg His Ala Gly Phe Val Gly
45                       50                       55                       60
Pro Ala Lys Leu Arg Gln
                  65
```

<210> 162
<211> 44
<212> PRT
<213> Homo sapiens

<400> 162
```
Met Ser Pro Arg Leu Glu Cys Ser Gly Ala Ile Leu Ala His Cys Asn
1                        5                       10                       15
Pro Arg Leu Pro Gly Ser Ser Tyr Ser Pro Ala Ser Ala Thr Trp Val
                  20                       25                       30
Arg Gly Ser Leu Glu Pro Gly Arg Leu Arg Leu Gln
              35                       40
```

<210> 163
<211> 314
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -58..-1

```
<400> 163
Met Gln Asn Val Ile Asn Thr Val Lys Gly Lys Ala Leu Glu Val Ala
            -55             -50             -45
Glu Tyr Leu Thr Pro Val Leu Lys Glu Ser Lys Phe Arg Glu Thr Gly
        -40             -35             -30
Val Ile Thr Pro Glu Glu Phe Val Ala Ala Gly Asp His Leu Val His
    -25             -20             -15
His Cys Pro Thr Trp Gln Trp Ala Thr Gly Glu Glu Leu Lys Val Lys
-10              -5               1               5
Ala Tyr Leu Pro Thr Gly Lys Gln Phe Leu Val Thr Lys Asn Val Pro
            10              15              20
Cys Tyr Lys Arg Cys Lys Gln Met Glu Tyr Ser Asp Glu Leu Glu Ala
        25              30              35
Ile Ile Glu Glu Asp Asp Gly Asp Gly Gly Trp Val Asp Thr Tyr His
    40              45              50
Asn Thr Gly Ile Thr Gly Ile Thr Glu Ala Val Lys Glu Ile Thr Leu
55              60              65              70
Glu Asn Lys Asp Asn Ile Arg Leu Gln Asp Cys Ser Ala Leu Cys Glu
        75              80              85
Glu Glu Glu Asp Glu Asp Glu Gly Glu Ala Ala Asp Met Glu Glu Tyr
        90              95              100
Glu Glu Ser Gly Leu Leu Glu Thr Asp Glu Ala Thr Leu Asp Thr Arg
        105             110             115
Lys Ile Val Glu Ala Cys Lys Ala Lys Thr Asp Ala Gly Gly Glu Asp
    120             125             130
Ala Ile Leu Gln Thr Arg Thr Tyr Asp Leu Tyr Ile Thr Tyr Asp Lys
135             140             145             150
Tyr Tyr Gln Thr Pro Arg Leu Trp Leu Phe Gly Tyr Asp Glu Gln Arg
            155             160             165
Gln Pro Leu Thr Val Glu His Met Tyr Glu Asp Ile Ser Gln Asp His
            170             175             180
Val Lys Lys Thr Val Thr Ile Glu Asn His Pro His Leu Pro Pro Pro
            185             190             195
Pro Met Cys Ser Val His Pro Cys Arg His Ala Glu Val Met Lys Lys
    200             205             210
Ile Ile Glu Thr Val Ala Glu Gly Gly Gly Glu Leu Gly Val His Met
215             220             225             230
Tyr Leu Leu Ile Phe Leu Lys Phe Val Gln Ala Val Ile Pro Thr Ile
            235             240             245
Glu Tyr Asp Tyr Thr Arg His Phe Thr Met
            250             255


<210> 164
<211> 89
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -80..-1

<400> 164
Met Arg Thr Arg Thr Thr Gly Asn Pro Arg Gly Leu His Asp Thr Phe
-80             -75             -70             -65
Pro Arg Arg Pro Arg Leu Gly Arg Cys Ser Asp Met Asp Thr Ala Arg
            -60             -55             -50
Thr Ser Cys Ser Asp Leu Leu Pro Trp Glu Gly Val Thr Glu Pro Ala
            -45             -40             -35
Leu Cys Gly Asp Gln Leu Gln Gly Thr Glu Gly Trp Leu Glu Ala Thr
```

```
                -30                    -25                    -20
Gln Leu Gly Arg Gly Leu Leu Ser Ala Cys Ala Pro Trp Gly Asp Gly
    -15                    -10                    -5
Ser Thr Gln Pro Val Pro Leu Cys Ser
1                    5
```

<210> 165
<211> 98
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -15..-1

<400> 165
```
Met Glu Ala Met Trp Leu Leu Cys Val Ala Leu Ala Val Leu Ala Trp
-15                    -10                    -5                    1
Gly Phe Leu Trp Val Trp Asp Ser Ser Glu Arg Met Lys Ser Arg Glu
         5                    10                    15
Gln Gly Gly Arg Leu Gly Ala Glu Ser Arg Thr Leu Leu Val Ile Ala
         20                    25                    30
His Pro Asp Asp Glu Ala Met Phe Phe Ala Pro Thr Val Leu Gly Leu
    35                    40                    45
Ala Arg Leu Arg His Trp Val Tyr Leu Leu Cys Phe Ser Ala Val Phe
50                    55                    60                    65
Arg Arg Glu Leu Ser Glu Tyr Thr Glu Gly Leu Thr Ser Glu Pro Leu
                  70                    75                    80
Thr Ala
```

<210> 166
<211> 92
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -36..-1

<400> 166
```
Met Leu Val Thr Gln Gly Leu Val Tyr Gln Gly Tyr Leu Ala Ala Asn
    -35                    -30                    -25
Ser Arg Phe Gly Ser Leu Pro Lys Val Ala Leu Ala Gly Leu Leu Gly
-20                    -15                    -10                    -5
Phe Gly Leu Gly Lys Val Ser Tyr Ile Gly Val Cys Gln Ser Lys Phe
                  1                    5                    10
His Phe Phe Glu Asp Gln Leu Arg Gly Ala Gly Phe Gly Pro Gln His
         15                    20                    25
Asn Arg His Cys Leu Leu Thr Cys Glu Glu Cys Lys Ile Lys His Gly
    30                    35                    40
Leu Ser Glu Lys Gly Asp Ser Gln Pro Ser Ala Ser
45                    50                    55
```

<210> 167
<211> 351
<212> PRT

<213> Homo sapiens

<220>
<221> SIGNAL
<222> -16..-1

<400> 167
Met Val Pro Phe Ile Tyr Leu Gln Ala His Phe Thr Leu Cys Ser Gly
    -15   -10   -5
Trp Ser Ser Thr Tyr Arg Asp Leu Arg Lys Gly Val Tyr Val Pro Tyr
1   5   10   15
Thr Gln Gly Lys Trp Glu Gly Glu Leu Gly Thr Asp Leu Val Ser Ile
            20   25   30
Pro His Gly Pro Asn Val Thr Val Arg Ala Asn Ile Ala Ala Ile Thr
        35   40   45
Glu Ser Asp Lys Phe Phe Ile Asn Gly Ser Asn Trp Glu Gly Ile Leu
    50   55   60
Gly Leu Ala Tyr Ala Glu Ile Ala Arg Pro Asp Asp Ser Pro Glu Pro
65   70   75   80
Phe Phe Asp Ser Leu Val Lys Gln Thr His Val Pro Asn Leu Phe Ser
            85   90   95
Leu Gln Leu Cys Gly Ala Gly Phe Pro Leu Asn Gln Ser Glu Val Leu
            100   105   110
Ala Ser Val Gly Gly Ser Met Ile Ile Gly Gly Ile Asp His Ser Leu
            115   120   125
Tyr Thr Gly Ser Leu Trp Tyr Thr Pro Ile Arg Arg Glu Trp Tyr Tyr
    130   135   140
Glu Val Ile Ile Val Arg Val Glu Ile Asn Gly Gln Asp Leu Lys Met
145   150   155   160
Asp Cys Lys Glu Tyr Asn Tyr Asp Lys Ser Ile Val Asp Ser Gly Thr
            165   170   175
Thr Asn Leu Arg Leu Pro Lys Lys Val Phe Glu Ala Ala Val Lys Ser
            180   185   190
Ile Lys Ala Ala Ser Ser Thr Glu Lys Phe Pro Asp Gly Phe Trp Leu
    195   200   205
Gly Glu Gln Leu Val Cys Trp Gln Ala Gly Thr Thr Pro Trp Asn Ile
    210   215   220
Phe Pro Val Ile Ser Leu Tyr Leu Met Gly Glu Val Thr Asn Gln Ser
225   230   235   240
Phe Arg Ile Thr Ile Leu Pro Gln Gln Tyr Leu Arg Pro Val Glu Asp
            245   250   255
Val Ala Thr Ser Gln Asp Asp Cys Tyr Lys Phe Ala Ile Ser Gln Ser
            260   265   270
Ser Thr Gly Thr Val Met Gly Ala Val Ile Met Glu Gly Phe Tyr Val
    275   280   285
Val Phe Asp Arg Ala Arg Lys Arg Ile Gly Phe Ala Val Ser Ala Cys
    290   295   300
His Val His Asp Glu Phe Arg Thr Ala Ala Val Glu Gly Pro Phe Cys
305   310   315   320
His Leu Gly His Gly Arg Leu Trp Leu Gln His Ser Thr Asp Arg
            325   330   335

<210> 168
<211> 138
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -47..-1

```
<400> 168
Met Glu Lys Phe Val Asp Pro Gly Asn His Asn Ser Gly Ile Asp Leu
        -45                 -40                 -35
Leu Arg Thr Tyr Leu Trp Arg Cys Gln Phe Leu Leu Pro Phe Val Ser
        -30                 -25                 -20
Leu Gly Leu Met Cys Phe Gly Ala Leu Ile Gly Leu Cys Ala Cys Ile
-15                 -10                 -5                      1
Cys Arg Ser Leu Tyr Pro Thr Ile Ala Thr Gly Ile Leu His Leu Leu
            5                   10.                 15
Ala Gly Leu Cys Thr Leu Gly Ser Val Ser Cys Tyr Val Ala Gly Ile
            20                  25                  30
Glu Leu Leu His Gln Lys Leu Glu Leu Pro Asp Asn Val Ser Gly Glu
        35                  40                  45
Phe Gly Trp Ser Phe Cys Leu Ala Cys Val Ser Ala Pro Leu Gln Phe
50                  55                  60                      65
Met Ala Ser Ala Leu Phe Ile Trp Ala Ala His Thr Asn Arg Arg Glu
                70                  75                  30
Tyr Thr Leu Met Lys Ala Tyr Arg Val Ala
                85                  90
```

```
<210> 169
<211> 101
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -73..-1

<400> 169
Met Asn Leu Glu Arg Val Ser Asn Glu Glu Lys Leu Asn Leu Cys Arg
            -70                 -65                 -60
Lys Tyr Tyr Leu Gly Gly Phe Ala Phe Leu Pro Phe Leu Trp Leu Val
            -55                 -50                 -45
Asn Ile Phe Trp Phe Tyr Arg Glu Ala Phe Leu Val Pro Ala Tyr Thr
        -40                 -35                 -30
Glu Gln Ser Gln Ile Lys Gly Tyr Val Trp Arg Ser Ala Val Gly Phe
-25                 -20                 -15                 -10
Leu Phe Trp Val Ile Val Leu Thr Ser Trp Ile Thr Ile Phe Gln Ile
                -5                  1                   5
Tyr Arg Pro Arg Trp Gly Ala Leu Gly Asp Tyr Leu Ser Phe Thr Ile
        10                  15                  20
Pro Leu Gly Thr Pro
    25
```

```
<210> 170
<211> 252
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -68..-1

<400> 170
Met Pro Glu Gly Pro Glu Leu His Leu Ala Ser Gln Phe Val Asn Glu
            -65                 -60                 -55
```

```
Ala Cys Arg Ala Leu Val Phe Gly Gly Cys Val Glu Lys Ser Ser Val
        -50                 -45                 -40
Ser Arg Asn Pro Glu Val Pro Phe Glu Ser Ser Ala Tyr Arg Ile Ser
    -35                 -30                 -25
Ala Ser Ala Arg Gly Lys Glu Leu Arg Leu Ile Leu Ser Pro Leu Pro
-20                 -15                 -10                 -5
Gly Ala Gln Pro Gln Gln Glu Pro Leu Ala Leu Val Phe Arg Phe Gly
                1               5                   10
Met Ser Gly Ser Phe Gln Leu Val Pro Arg Glu Glu Leu Pro Arg His
        15                  20                  25
Ala His Leu Arg Phe Tyr Thr Ala Pro Pro Gly Pro Arg Leu Ala Leu
        30                  35                  40
Cys Phe Val Asp Ile Arg Arg Phe Gly Arg Trp Asp Leu Gly Gly Lys
45                  50                  55                  60
Trp Gln Pro Gly Arg Gly Pro Cys Val Leu Gln Glu Tyr Gln Gln Phe
                65                  70                  75
Arg Glu Asn Val Leu Arg Asn Leu Ala Asp Lys Ala Phe Asp Arg Pro
            80                  85                  90
Ile Cys Glu Ala Leu Leu Asp Gln Arg Phe Phe Asn Gly Ile Gly Asn
            95                  100                 105
Tyr Leu Arg Ala Glu Ile Leu Tyr Arg Leu Lys Ile Pro Pro Phe Glu
110                 115                 120
Lys Ala Arg Ser Val Leu Glu Ala Leu Gln Gln His Arg Pro Ser Pro
125                 130                 135                 140
Glu Leu Thr Leu Ser Gln Lys Ile Arg Thr Lys Leu Gln Asn Ser Asp
                145                 150                 155
Leu Leu Glu Leu Cys His Ser Val Pro Lys Glu Val Val Gln Leu Gly
                160                 165                 170
Glu Ala Lys Asp Gly Ser Asn Leu Cys Phe Ser Lys
        175                 180
```

<210> 171
<211> 350
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -68..-1

<400> 171
```
Met Pro Glu Gly Pro Glu Leu His Leu Ala Ser Gln Phe Val Asn Glu
            -65                 -60                 -55
Ala Cys Arg Ala Leu Val Phe Gly Gly Cys Val Glu Lys Ser Ser Val
        -50                 -45                 -40
Ser Arg Asn Pro Glu Val Pro Phe Glu Ser Ser Ala Tyr Arg Ile Ser
    -35                 -30                 -25
Ala Ser Ala Arg Gly Lys Glu Leu Arg Leu Ile Leu Ser Pro Leu Pro
-20                 -15                 -10                 -5
Gly Ala Gln Pro Gln Gln Glu Pro Leu Ala Leu Val Phe Arg Phe Gly
                1               5                   10
Met Ser Gly Ser Phe Gln Leu Val Pro Arg Glu Glu Leu Pro Arg His
        15                  20                  25
Ala His Leu Arg Phe Tyr Thr Ala Pro Pro Gly Pro Arg Leu Ala Leu
        30                  35                  40
Cys Phe Val Asp Ile Arg Arg Phe Gly Arg Trp Asp Leu Gly Gly Lys
45                  50                  55                  60
Trp Gln Pro Gly Arg Gly Pro Cys Val Leu Gln Glu Tyr Gln Gln Phe
                65                  70                  75
Arg Leu Lys Ile Pro Pro Phe Glu Lys Ala Arg Ser Val Leu Glu Ala
```

```
                    80                    85                    90
Leu Gln Gln His Arg Pro Ser Pro Glu Leu Thr Leu Ser Gln Lys Ile
            95                   100                   105
Arg Thr Lys Leu Gln Asn Pro Asp Leu Leu Glu Leu Cys His Ser Val
    110                   115                   120
Pro Lys Glu Val Asp Gln Leu Gly Gly Arg Gly Tyr Gly Ser Glu Ser
125                   130                   135                   140
Gly Glu Glu Asp Phe Ala Ala Phe Arg Ala Trp Leu Arg Cys Tyr Gly
                145                   150                   155
Met Pro Gly Met Ser Ser Leu Gln Asp Arg His Gly Arg Thr Ile Trp
            160                   165                   170
Phe Gln Gly Asp Pro Gly Pro Leu Ala Pro Lys Gly Arg Lys Ser Arg
            175                   180                   185
Lys Lys Lys Ser Lys Ala Thr Gln Leu Ser Pro Glu Asp Arg Val Glu
            190                   195                   200
Asp Ala Leu Pro Pro Ser Lys Ala Pro Ser Lys Thr Arg Arg Ala Lys
205                   210                   215                   220
Arg Asp Leu Pro Lys Arg Thr Ala Thr Gln Arg Pro Glu Gly Thr Ser
                225                   230                   235
Leu Gln Gln Asp Pro Glu Ala Pro Thr Val Pro Lys Lys Gly Arg Arg
            240                   245                   250
Lys Gly Arg Gln Ala Ala Ser Gly His Cys Arg Pro Arg Lys Val Lys
            255                   260                   265
Ala Asp Ile Pro Ser Leu Glu Pro Glu Gly Thr Ser Ala Ser
    270                   275                   280


<210> 172
<211> 390
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -68..-1

<400> 172
Met Pro Glu Gly Pro Glu Leu His Leu Ala Ser Gln Phe Val Asn Glu
                -65                   -60                   -55
Ala Cys Arg Ala Leu Val Phe Gly Gly Cys Val Glu Lys Ser Ser Val
        -50                   -45                   -40
Ser Arg Asn Pro Glu Val Pro Phe Glu Ser Ser Ala Tyr Arg Ile Ser
        -35                   -30                   -25
Ala Ser Ala Arg Gly Lys Glu Leu Arg Leu Ile Leu Ser Pro Leu Pro
-20                   -15                   -10                   -5
Gly Ala Gln Pro Gln Gln Glu Pro Leu Ala Leu Val Phe Arg Phe Gly
                1                     5                     10
Met Ser Gly Ser Phe Gln Leu Val Pro Arg Glu Glu Leu Pro Arg His
            15                    20                    25
Ala His Leu Arg Phe Tyr Thr Ala Pro Pro Gly Pro Arg Leu Ala Leu
    30                    35                    40
Cys Phe Val Asp Ile Arg Arg Phe Gly Arg Trp Asp Leu Gly Gly Lys
45                    50                    55                    60
Trp Gln Pro Gly Arg Gly Pro Cys Val Leu Gln Glu Tyr Gln Gln Phe
                65                    70                    75
Arg Glu Asn Val Leu Arg Asn Leu Ala Asp Lys Ala Phe Asp Arg Pro
            80                    85                    90
Ile Cys Glu Ala Leu Leu Asp Gln Arg Phe Phe Asn Gly Ile Gly Asn
            95                    100                   105
Tyr Leu Arg Ala Glu Ile Leu Tyr Arg Leu Lys Ile Pro Pro Phe Glu
    110                   115                   120
```

```
Lys Ala Arg Ser Val Leu Glu Ala Leu Gln Gln His Arg Pro Ser Pro
125             130             135             140
Glu Leu Thr Leu Ser Gln Lys Ile Arg Thr Lys Leu Gln Asn Pro Asp
            145             150             155
Leu Leu Glu Leu Cys His Ser Val Pro Lys Glu Val Val Gln Leu Gly
        160             165             170
Gly Arg Gly Tyr Gly Ser Glu Ser Gly Glu Glu Asp Phe Ala Ala Phe
    175             180             185
Arg Ala Trp Leu Arg Cys Tyr Gly Met Pro Gly Met Ser Ser Leu Gln
    190             195             200
Asp Arg His Gly Arg Thr Ile Trp Phe Gln Gly Asp Pro Gly Pro Leu
205             210             215             220
Ala Pro Lys Gly Arg Lys Ser Arg Lys Lys Lys Ser Lys Ala Thr Gln
            225             230             235
Leu Ser Pro Glu Asp Arg Val Glu Asp Ala Leu Pro Pro Ser Lys Ala
            240             245             250
Pro Ser Arg Thr Arg Arg Ala Lys Arg Asp Leu Pro Lys Arg Thr Ala
            255             260             265
Thr Gln Arg Pro Glu Gly Thr Ser Leu Gln Gln Asp Pro Glu Ala Pro
    270             275             280
Thr Val Pro Lys Lys Gly Arg Arg Lys Gly Arg Gln Ala Ala Ser Gly
285             290             295             300
His Cys Arg Pro Arg Lys Val Lys Ala Asp Ile Pro Ser Leu Glu Pro
            305             310             315
Glu Gly Thr Ser Ala Ser
            320
```

```
<210> 173
<211> 190
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -82..-1

<400> 173
Met Tyr Val Trp Pro Cys Ala Val Val Leu Ala Gln Tyr Leu Trp Phe
        -80             -75             -70
His Arg Arg Ser Leu Pro Gly Lys Ala Ile Leu Glu Ile Gly Ala Gly
    -65             -60             -55
Val Ser Leu Pro Gly Ile Leu Thr Ala Lys Cys Gly Ala Glu Val Ile
-50             -45             -40                 -35
Leu Ser Asp Ser Ser Glu Leu Pro His Cys Leu Glu Val Cys Arg Gln
            -30             -25                 -20
Ser Cys Gln Met Asn Asn Leu Pro His Leu Gln Val Val Gly Leu Thr
            -15             -10             -5
Trp Gly His Ile Ser Trp Asp Leu Leu Ala Leu Pro Pro Gln Asp Ile
    1               5               10
Ile Leu Ala Ser Asp Val Phe Phe Glu Pro Glu Asp Phe Glu Asp Ile
15              20              25              30
Leu Ala Thr Ile Tyr Phe Leu Met His Lys Asn Pro Lys Val Gln Leu
            35              40              45
Trp Ser Thr Tyr Gln Val Arg Ser Ala Asp Trp Ser Leu Glu Ala Leu
        50              55              60
Leu Tyr Lys Trp Asp Met Lys Cys Val His Ile Pro Leu Glu Ser Phe
        65              70              75
Asp Ala Asp Lys Glu Asp Ile Ala Glu Ser Thr Leu Pro Gly Arg His
    80              85              90
Thr Val Glu Met Leu Val Ile Ser Phe Ala Lys Asp Ser Leu
```

95   100   105

<210> 174
<211> 285
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -232..-1

<400> 174
Met Gly Cys Val Phe Gln Ser Thr Glu Asp Lys Arg Ile Phe Lys Ile
        -230   -225   -220
Asp Trp Thr Leu Ser Pro Gly Glu His Ala Lys Asp Glu Tyr Val Leu
    -215   -210   -205
Tyr Tyr Tyr Ser Asn Leu Ser Val Pro Ile Gly Arg Phe Gln Asn Arg
-200   -195   -190   -185
Val His Leu Met Gly Asp Asn Leu Cys Asn Asp Gly Ser Leu Leu Leu
                -180   -175   -170
Gln Asp Val Gln Glu Ala Asp Gln Gly Thr Tyr Ile Cys Glu Ile Arg
            -165   -160   -155
Leu Lys Gly Glu Ser Gln Val Phe Lys Lys Ala Val Val Leu His Val
            -150   -145   -140
Leu Pro Glu Glu Pro Lys Glu Leu Met Val His Val Gly Gly Leu Ile
    -135   -130   -125
Gln Met Gly Cys Val Phe Gln Ser Thr Glu Val Lys His Val Thr Lys
-120   -115   -110   -105
Val Glu Trp Ile Phe Ser Gly Arg Arg Ala Lys Glu Glu Ile Val Phe
                -100   -95   -90
Arg Tyr Tyr His Lys Leu Arg Met Ser Ala Glu Tyr Ser Gln Ser Trp
            -85   -80   -75
Gly His Phe Gln Asn Arg Val Asn Leu Val Gly Asp Ile Phe Arg Asn
        -70   -65   -60
Asp Gly Ser Ile Met Leu Gln Gly Val Arg Glu Ser Asp Gly Gly Asn
    -55   -50   -45
Tyr Thr Cys Ser Ile His Leu Gly Asn Leu Val Phe Lys Lys Thr Ile
-40   -35   -30   -25
Val Leu His Val Ser Pro Glu Glu Pro Arg Thr Leu Val Thr Pro Ala
                -20   -15   -10
Ala Leu Arg Pro Leu Val Leu Gly Gly Asn Gln Leu Val Ile Ile Val
            -5   1   5
Gly Ile Val Cys Ala Thr Ile Leu Leu Leu Pro Val Leu Ile Leu Ile
        10   15   20
Val Lys Lys Thr Cys Gly Asn Lys Ser Ser Val Asn Ser Thr Val Leu
25   30   35   40
Val Lys Asn Thr Lys Lys Thr Asn Pro Lys Lys Lys Lys
                45   50

<210> 175
<211> 153
<212> PRT
<213> Homo sapiens

<400> 175
Met Gly Cys Val Phe Gln Ser Thr Val Asp Lys Cys Ile Phe Lys Ile
1   5   10   15
Asp Trp Thr Leu Ser Pro Gly Glu His Ala Lys Asp Glu Tyr Val Leu

```
                     20                    25                    30
Tyr Tyr Tyr Ser Asn Leu Ser Val Pro Ile Gly Arg Phe Gln Asn Arg
         35                    40                    45
Val His Leu Met Gly Asp Ile Leu Cys Asn Asp Gly Ser Leu Leu Leu
   50                    55                    60
Gln Asp Val Gln Glu Ala Asp Gln Gly Thr Tyr Ile Cys Glu Ile Arg
65                    70                    75                    80
Leu Lys Gly Glu Ser Gln Val Phe Lys Lys Ala Val Val Leu His Val
                  85                    90                    95
Leu Pro Glu Glu Pro Lys Glu Leu Met Val His Val Gly Gly Leu Ile
               100                   105                   110
Gln Met Gly Cys Val Phe Gln Ser Thr Glu Val Lys His Val Thr Lys
         115                   120                   125
Val Glu Trp Ile Phe Ser Gly Arg Arg Ala Lys Val Thr Arg Arg Lys
   130                   135                   140
His His Cys Val Arg Glu Gly Ser Gly
145                   150
```

```
<210> 176
<211> 49
<212> PRT
<213> Homo sapiens

<400> 176
Met Leu Xaa Gly Asp His Arg Ala Leu Leu Leu Lys Ile Trp Leu Leu
1             5                     10                    15
Gln Arg Pro Glu Ser Gln Glu Gly Leu Leu Pro Gly Arg Leu Val Val
            20                    25                    30
Met Glu Arg Arg Val Lys Met Thr Ser Cys Pro Ser Cys Pro Arg Phe
         35                    40                    45
Cys
```

```
<210> 177
<211> 99
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -24..-1

<400> 177
Met Lys Ser Ala Lys Leu Gly Phe Leu Leu Arg Phe Phe Ile Phe Cys
            -20                   -15                   -10
Ser Leu Asn Thr Leu Leu Leu Gly Gly Val Asn Lys Ile Ala Glu Lys
            -5                    1                     5
Ile Cys Gly Asp Leu Lys Asp Pro Cys Lys Leu Asp Met Asn Phe Gly
   10                    15                    20
Ser Cys Tyr Glu Val His Phe Arg Tyr Phe Tyr Asn Arg Thr Ser Lys
25                    30                    35                    40
Arg Cys Glu Thr Phe Val Phe Ser Gly Cys Asn Gly Asn Leu Asn Asn
                  45                    50                    55
Phe Lys Leu Lys Ile Glu Arg Glu Val Ala Cys Val Ala Lys Tyr Lys
               60                    65                    70
Pro Pro Arg
         75
```

```
<210> 178
<211> 95
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -37..-1

<400> 178
Met Ala Ser Pro Ala Val Asn Arg Trp Lys Arg Pro Arg Leu Lys Pro
        -35                   -30                   -25
Val Trp Pro Arg Arg Leu Glu Ser Trp Leu Leu Leu Asp Ala Leu Leu
        -20                   -15                   -10
Arg Leu Gly Asp Thr Lys Lys Lys Arg Gln Pro Glu Ala Ala Thr Lys
    -5                    1                     5                     10
Ser Cys Val Arg Ser Ser Cys Gly Gly Pro Ser Gly Asp Gly Pro Pro
            15                    20                    25
Pro Cys Leu Gln Gln Pro Asp Pro Arg Ala Leu Ser Gln Ala Phe Ser
        30                    35                    40
Arg Ser Phe Pro Leu Phe Pro Ser Leu Ala Gly Lys Ser Met Ile
    45                    50                    55


<210> 179
<211> 121
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -23..-1

<400> 179
Met Met Leu Pro Gln Trp Leu Leu Leu Leu Phe Leu Leu Phe Phe Phe
            -20                   -15                   -10
Leu Phe Leu Leu Thr Arg Gly Ser Leu Ser Pro Thr Lys Tyr Asn Leu
    -5                    1                     5
Leu Glu Leu Lys Glu Ser Cys Ile Arg Asn Gln Asp Cys Glu Thr Gly
10                    15                    20                    25
Cys Cys Gln Arg Ala Pro Asp Asn Cys Glu Ser His Cys Ala Glu Lys
            30                    35                    40
Gly Ser Glu Gly Ser Leu Cys Gln Thr Gln Val Phe Phe Gly Gln Tyr
            45                    50                    55
Arg Ala Cys Pro Cys Leu Arg Asn Leu Thr Cys Ile Tyr Ser Lys Asn
        60                    65                    70
Glu Lys Trp Leu Ser Ile Ala Tyr Gly Arg Cys Gln Lys Ile Gly Arg
    75                    80                    85
Gln Lys Leu Ala Lys Lys Met Phe Phe
90                    95


<210> 180
<211> 59
<212> PRT
<213> Homo sapiens

<400> 180
Met Ile Leu Cys Phe Leu Leu Pro His His Arg Leu Gln Glu Ala Arg
```

```
1               5                   10                  15
Gln Ile Gln Val Leu Lys Met Leu Pro Arg Glu Lys Leu Arg Arg Arg
            20                  25                  30
Glu Glu Arg Lys Gln Ile Asn Gly Lys Lys Glu Arg Thr Lys Tyr Glu
        35                  40                  45
Thr Pro Arg Lys Arg Glu Gly Lys Lys Lys Lys
        50                  55
```

<210> 181
<211> 86
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -14..-1

<400> 181
```
Met Val Ala Leu Asn Leu Ile Leu Val Pro Cys Cys Ala Ala Trp Cys
                -10                 -5                  1
Asp Pro Arg Arg Ile His Ser Gln Asp Asp Val Pro Arg Ser Ser Ala
            5                   10                  15
Ala Asp Thr Gly Ser Ala Met Gln Arg Arg Glu Ala Trp Ala Gly Trp
        20                  25                  30
Arg Arg Ser Gln Pro Phe Ser Val Gly Leu Pro Ser Ala Glu Arg Leu
35                  40                  45                  50
Glu Asn Gln Pro Gly Lys Leu Ser Trp Arg Ser Leu Val Gly Glu Gly
                55                  60                  65
Tyr Arg Ile Cys Asp Leu
            70
```

<210> 182
<211> 165
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -58..-1

<400> 182
```
Met Thr Arg Leu Cys Leu Pro Arg Pro Glu Ala Arg Glu Asp Pro Ile
            -55                 -50                 -45
Pro Val Pro Pro Arg Gly Leu Gly Ala Gly Glu Gly Ser Gly Ser Pro
            -40                 -35                 -30
Val Arg Pro Pro Val Ser Thr Trp Gly Pro Ser Trp Ala Gln Leu Leu
            -25                 -20                 -15
Asp Ser Val Leu Trp Leu Gly Ala Leu Gly Leu Thr Ile Gln Ala Val
-10                 -5                  1                   5
Phe Ser Thr Thr Gly Pro Ala Leu Leu Leu Leu Leu Val Ser Phe Leu
            10                  15                  20
Thr Phe Asp Leu Leu His Arg Pro Ala Gly His Thr Leu Pro Gln Arg
            25                  30                  35
Lys Leu Leu Thr Arg Gly Gln Ser Gln Gly Ala Gly Glu Gly Pro Gly
            40                  45                  50
Gln Gln Glu Ala Leu Leu Leu Gln Met Gly Thr Val Ser Gly Gln Leu
55                  60                  65                  70
Ser Leu Gln Asp Ala Leu Leu Leu Leu Leu Met Gly Leu Gly Pro Leu
```

```
                          75                      80                      85
Leu Arg Ala Cys Gly Met Pro Leu Thr Leu Leu Gly Leu Ala Phe Cys
                  90                      95                     100
Leu His Pro Trp Ala
          105




<210> 183
<211> 80
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -35..-1

<400> 183
Met Pro Phe Gln Phe Gly Thr Gln Pro Arg Arg Phe Pro Val Glu Gly
-35                     -30                     -25                     -20
Gly Asp Ser Ser Ile Glu Leu Glu Pro Gly Leu Ser Ser Ser Ala Ala
                -15                     -10                     -5
Cys Asn Gly Lys Glu Met Ser Pro Thr Arg Gln Leu Arg Arg Cys Pro
              1                   5                      10
Gly Ser His Cys Leu Thr Ile Thr Asp Val Pro Val Thr Val Tyr Ala
        15                      20                      25
Thr Thr Arg Lys Pro Pro Ala Gln Ser Ser Lys Glu Met His Pro Lys
30                      35                      40                      45




<210> 184
<211> 73
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 184
Met Ala Pro Gln Thr Leu Leu Pro Val Leu Val Leu Cys Val Leu Leu
        -20                     -15                     -10
Leu Gln Ala Gln Gly Gly Tyr Arg Asp Lys Met Arg Met Gln Arg Ile
-5                  1               5                       10
Lys Val Cys Glu Lys Arg Pro Ser Ile Asp Leu Cys Ile His His Cys
              15                  20                      25
Ser Cys Phe Gln Lys Cys Glu Thr Asn Lys Ile Cys Cys Ser Ala Phe
          30                      35                      40
Cys Gly Asn Ile Cys Met Ser Ile Leu
        45                      50




<210> 185
<211> 98
<212> PRT
<213> Homo sapiens

<400> 185
Met Leu Gly Ala Glu Thr Glu Glu Lys Leu Phe Asp Ala Pro Leu Ser
1                   5                       10                      15
```

```
Ile Ser Lys Arg Glu Gln Leu Glu Gln Gln Val Pro Glu Asn Tyr Phe
            20              25              30
Tyr Val Pro Asp Leu Gly Gln Val Pro Glu Ile Asp Val Pro Ser Tyr
        35              40              45
Leu Pro Asp Leu Pro Gly Ile Ala Asn Asp Leu Met Tyr Ile Ala Asp
    50              55              60
Leu Gly Pro Gly Ile Ala Pro Ser Ala Pro Gly Thr Ile Pro Glu Leu
65              70              75              80
Pro Thr Phe His Thr Glu Val Ala Glu Pro Leu Lys Thr Tyr Lys Met
            85              90              95
Gly Tyr
```

```
<210> 186
<211> 112
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 186
Met Glu Ser Arg Val Leu Leu Arg Thr Phe Cys Leu Ile Phe Gly Leu
    -20             -15             -10
Gly Ala Val Trp Gly Leu Gly Val Asp Pro Ser Leu Gln Ile Asp Val
-5              1               5               10
Leu Thr Glu Leu Glu Leu Gly Glu Ser Thr Thr Gly Val Arg Gln Val
            15              20              25
Pro Gly Leu His Asn Gly Thr Lys Ala Phe Leu Phe Gln Asp Thr Pro
        30              35              40
Arg Ser Ile Lys Ala Ser Thr Ala Thr Ala Glu Gln Phe Phe Gln Lys
    45              50              55
Leu Arg Asn Lys His Glu Phe Thr Ile Leu Val Thr Leu Lys Gln Thr
60              65              70              75
His Leu Asn Ser Gly Val Ile Leu Ser Ile His His Leu Asp His Arg
            80              85              90
```

```
<210> 187
<211> 70
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -44..-1

<400> 187
Met Cys Cys Tyr Cys Arg Ile Phe Cys Leu Arg Cys Thr Tyr Phe Pro
            -40             -35             -30
Val His Cys Gly Met Cys Asn Leu Arg Tyr Phe Glu Phe Ser Thr Phe
        -25             -20             -15
Leu Leu Ser Leu Ser Leu Ile Thr Tyr Cys Phe Trp Asp Pro Pro His
    -10             -5              1
Arg Gly Ser His Ser Ser Leu Ser Leu Glu His Thr Pro Leu Asp Phe Leu
5               10              15              20
Glu Trp Gly Leu Leu Arg
            25
```

<210> 188
<211> 92
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -13..-1

<400> 188
Met Leu Phe Ser Leu Ser Leu Leu Ser Asn Leu Asn Gln Ile Gly Ser
          -10   -5   1
Ser His Leu Asp Arg Pro His Ile Pro Gly Gln Ser Ala Gln Leu Phe
    5   10   15
Ile Tyr Gln Met Ser Ser Gln Gln Leu Gln Gln Gln Pro Ser Ala Asn
20   25   30   35
Lys Lys Ala Gly Lys Ile His Asn Thr Pro Phe Ala Asn Gln Leu Asn
              40   45   50
Pro Thr Gln His Leu Ala Lys Pro Phe Gln Gln Ile Leu Pro Gly Arg
              55   60   65
Gln Ser Gly Ser Leu Thr Ser Pro Phe Leu Ala Cys
          70   75

<210> 189
<211> 207
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -42..-1

<400> 189
Met His Ile Leu Gln Leu Leu Thr Thr Val Asp Asp Gly Ile Gln Ala
          -40   -35   -30
Ile Val His Cys Pro Asp Thr Gly Lys Asp Ile Trp Asn Leu Leu Phe
    -25   -20   -15
Asp Leu Val Cys His Glu Phe Cys Gln Ser Asp Asp Pro Pro Ile Ile
-10   -5   1   5
Leu Gln Glu Gln Lys Thr Val Leu Ala Ser Val Phe Ser Val Leu Ser
              10   15   20
Ala Ile Tyr Ala Ser Gln Thr Glu Gln Glu Tyr Leu Lys Ile Glu Lys
          25   30   35
Val Asp Leu Pro Leu Ile Asp Ser Leu Ile Arg Val Leu Gln Asn Met
    40   45   50
Glu Gln Cys Gln Lys Lys Pro Glu Asn Ser Ala Glu Ser Asn Thr Glu
55   60   65   70
Glu Thr Lys Arg Thr Asp Leu Thr Gln Asp Asp Leu His Leu Lys Ile
              75   80   85
Leu Lys Asp Ile Leu Cys Glu Phe Leu Ser Asn Ile Phe Gln Ala Leu
          90   95   100
Thr Lys Glu Thr Val Ala Gln Gly Val Lys Glu Gly Gln Leu Ser Lys
      105   110   115
Gln Lys Cys Ser Ser Ala Phe Gln Asn Leu Leu Pro Phe Tyr Ser Pro
      120   125   130
Val Val Glu Asp Phe Ile Lys Ile Leu Arg Glu Val Asp Lys Ala Leu
135   140   145   150
Ala Asp Asp Leu Glu Lys Asn Phe Pro Ser Leu Lys Val Gln Thr

155   160   165

<210> 190
<211> 201
<212> PRT
<213> Homo sapiens

<400> 190
Met Gln Val Ala Leu Lys Glu Asp Leu Asp Ala Leu Lys Glu Lys Phe
1   5   10   15
Arg Thr Met Glu Ser Asn Gln Lys Ser Ser Phe Gln Glu Ile Pro Lys
            20   25   30
Leu Asn Glu Glu Leu Leu Ser Lys Gln Lys Gln Leu Glu Lys Ile Glu
        35   40   45
Ser Gly Glu Met Gly Leu Asn Lys Val Trp Ile Asn Ile Thr Glu Met
    50   55   60
Asn Lys Gln Ile Ser Leu Leu Thr Ser Ala Val Asn His Leu Lys Ala
65   70   75   80
Asn Val Lys Ser Ala Ala Asp Leu Ile Ser Leu Pro Thr Thr Val Glu
                85   90   95
Gly Leu Gln Lys Ser Val Ala Ser Ile Gly Asn Thr Leu Asn Ser Val
            100   105   110
His Leu Ala Val Glu Ala Leu Gln Lys Thr Val Asp Glu His Lys Lys
        115   120   125
Thr Met Glu Leu Leu Gln Ser Asp Met Asn Gln His Phe Leu Lys Glu
    130   135   140
Thr Pro Gly Ser Asn Gln Ile Ile Pro Ser Pro Ser Ala Thr Ser Glu
145   150   155   160
Leu Asp Asn Lys Thr His Ser Glu Asn Leu Lys Gln Met Gly Asp Arg
                165   170   175
Ser Ala Thr Leu Lys Arg Gln Ser Leu Asp Gln Val Thr Asn Arg Thr
                180   185   190
Asp Thr Val Lys Ile Gln Lys Lys Lys
            195   200

<210> 191
<211> 379
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -37..-1

<400> 191
Met Pro His Ser Ser Leu His Pro Ser Ile Pro Cys Pro Arg Gly His
        -35   -30   -25
Gly Ala Gln Lys Ala Ala Leu Val Leu Leu Ser Ala Cys Leu Val Thr
    -20   -15   -10
Leu Trp Gly Leu Gly Glu Pro Pro Glu His Thr Leu Arg Tyr Leu Val
-5   1   5   10
Leu His Leu Ala Ser Leu Gln Leu Gly Leu Leu Leu Asn Gly Val Cys
        15   20   25
Ser Leu Ala Glu Glu Leu Arg His Ile His Ser Arg Tyr Arg Gly Ser
        30   35   40
Tyr Trp Arg Thr Val Arg Ala Cys Leu Gly Cys Pro Leu Arg Arg Gly
    45   50   55
Ala Leu Leu Leu Leu Ser Ile Tyr Phe Tyr Tyr Ser Leu Pro Asn Ala

```
60                    65                    70                    75
Val Gly Pro Pro Phe Thr Trp Met Leu Ala Leu Leu Gly Leu Ser Gln
              80                    85                    90
Ala Leu Asn Ile Leu Leu Gly Leu Lys Gly Leu Ala Pro Ala Glu Ile
              95                    100                   105
Ser Ala Val Cys Glu Lys Gly Asn Phe Asn Val Ala His Gly Leu Ala
         110                   115                   120
Trp Ser Tyr Tyr Ile Gly Tyr Leu Arg Leu Ile Leu Pro Glu Leu Gln
         125              130                   135
Ala Arg Ile Arg Thr Tyr Asn Gln His Tyr Asn Asn Leu Leu Arg Gly
140                   145                   150                   155
Ala Val Ser Gln Arg Leu Tyr Ile Leu Leu Pro Leu Asp Cys Gly Val
              160                   165                   170
Pro Asp Asn Leu Ser Met Ala Asp Pro Asn Ile Arg Phe Leu Asp Lys
         175              180                   185
Leu Pro Gln Gln Thr Gly Asp Arg Ala Gly Ile Lys Asp Arg Val Tyr
         190                   195                   200
Ser Asn Ser Ile Tyr Glu Leu Leu Glu Asn Gly Gln Arg Ala Gly Thr
         205                   210                   215
Cys Val Leu Glu Tyr Ala Thr Pro Leu Gln Thr Leu Phe Ala Met Ser
220              225                   230                   235
Gln Tyr Ser Gln Ala Gly Phe Ser Arg Glu Asp Arg Leu Glu Gln Ala
         240                   245                   250
Lys Leu Phe Cys Arg Thr Leu Glu Asp Ile Leu Ala Asp Ala Pro Glu
         255                   260                   265
Ser Gln Asn Asn Cys Arg Leu Ile Ala Tyr Gln Glu Pro Ala Asp Asp
         270                   275                   280
Ser Ser Phe Ser Leu Ser Gln Glu Val Leu Arg His Leu Arg Gln Glu
         285                   290                   295
Glu Lys Glu Glu Val Thr Val Gly Ser Leu Lys Thr Ser Ala Val Pro
300                   305                   310                   315
Ser Thr Ser Thr Met Ser Gln Glu Pro Glu Leu Leu Leu Ser Gly Met
              320                   325                   330
Gly Lys Pro Leu Pro Leu Arg Thr Asp Phe Ser
              335                   340
```

```
<210> 192
<211> 112
<212> PRT
<213> Homo sapiens

<400> 192
Met Pro Ser Glu Gly Arg Cys Trp Glu Thr Leu Lys Ala Leu Arg Ser
1              5                    10                    15
Ser Asp Lys Gly Arg Leu Cys Tyr Tyr Arg Asp Trp Leu Leu Arg Arg
              20                   25                    30
Glu Asp Val Leu Glu Glu Cys Met Ser Leu Pro Lys Leu Ser Ser Tyr
              35                   40                    45
Ser Gly Trp Val Val Glu His Val Leu Pro His Met Gln Glu Asn Gln
         50                   55                    60
Pro Leu Ser Glu Thr Ser Pro Ser Ser Thr Ser Ala Ser Ala Leu Asp
65                   70                   75                    80
Gln Pro Ser Phe Val Pro Lys Ser Pro Asp Ala Ser Ser Ala Phe Ser
              85                   90                    95
Pro Ala Ser Pro Ala Thr Pro Asn Gly Thr Lys Gly Lys Lys Lys Lys
              100                  105                   110
```

```
<210> 193
```

<211> 43
<212> PRT
<213> Homo sapiens

<400> 193
Ser Leu Pro Gln Ala Leu Trp Phe Gln Phe Phe Tyr His Ser Gly Ser
1   5   10   15
Ser Leu Glu Ser Pro Gly Met Leu Asn Gly Pro Phe Gln His Arg Asn
            20   25   30
Ser Arg Ile Met Thr His Arg Ser Ala Glu Lys
        35   40

<210> 194
<211> 51
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -16..-1

<400> 194
Met Leu Arg Ile Ala Leu Thr Leu Ile Pro Ser Met Leu Ser Arg Ala
    -15   -10   -5
Ala Gly Trp Cys Trp Tyr Lys Glu Pro Thr Gln Gln Phe Ser Tyr Leu
1   5   10   15
Cys Leu Pro Cys Leu Ser Trp Asn Lys Lys Gly Asn Val Leu Gln Leu
            20   25   30
Pro Asn Phe
        35

<210> 195
<211> 244
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -18..-1

<400> 195
Met Ala Asn Pro Lys Leu Leu Gly Leu Glu Leu Ser Glu Ala Glu Ala
            -15   -10   -5
Ile Gly Ala Asp Ser Ala Arg Phe Glu Glu Leu Leu Leu Gln Ala Ser
        1   5   10
Lys Glu Leu Gln Gln Ala Gln Thr Thr Arg Pro Glu Ser Thr Gln Ile
15   20   25   30
Gln Pro Gln Pro Gly Phe Cys Ile Lys Thr Asn Ser Ser Glu Gly Lys
            35   40   45
Val Phe Ile Asn Ile Cys His Ser Pro Ser Ile Pro Pro Pro Ala Asp
            50   55   60
Val Thr Glu Glu Glu Leu Leu Gln Met Leu Glu Glu Asp Gln Ala Gly
        65   70   75
Phe Arg Ile Pro Met Ser Leu Gly Glu Pro His Ala Glu Leu Asp Ala
        80   85   90
Lys Gly Gln Gly Cys Thr Ala Tyr Asp Val Ala Val Asn Ser Asp Phe
95   100   105   110
Tyr Arg Arg Met Gln Asn Ser Asp Phe Leu Arg Glu Leu Val Ile Thr

```
                      115                     120                     125
Ile Ala Arg Glu Gly Leu Glu Asp Ile Tyr Asn Leu Gln Leu Asn Pro
              130                     135                     140
Glu Trp Arg Met Met Lys Asn Arg Pro Phe Met Gly Ser Ile Ser Gln
          145                     150                     155
Gln Asn Ile Arg Ser Glu Gln Arg Pro Arg Ile Gln Glu Leu Gly Asp
      160                     165                     170
Leu Tyr Thr Pro Ala Pro Gly Arg Ala Glu Ser Gly Pro Glu Lys Pro
  175                     180                     185                     190
His Leu Asn Leu Trp Leu Glu Ala Pro Asp Leu Leu Leu Ala Glu Val
                  195                     200                     205
Asp Leu Pro Lys Leu Asp Gly Ala Leu Gly Leu Ser Leu Glu Ile Gly
              210                     215                     220
Arg Thr Ala Trp
          225
```

<210> 196
<211> 353
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -34..-1

<400> 196
```
Met Glu Arg Gly Leu Lys Ser Ala Asp Pro Arg Asp Gly Thr Gly Tyr
              -30                     -25                     -20
Thr Gly Trp Ala Gly Ile Ala Val Leu Tyr Leu His Leu Tyr Asp Val
          -15                     -10                     -5
Phe Gly Asp Pro Ala Tyr Leu Gln Leu Ala His Gly Tyr Val Lys Gln
          1                   5                   10
Ser Leu Asn Cys Leu Thr Lys Arg Ser Ile Thr Phe Leu Cys Gly Asp
15                  20                  25                  30
Ala Gly Pro Leu Ala Val Ala Ala Val Leu Tyr His Lys Met Asn Asn
              35                  40                  45
Glu Lys Gln Ala Glu Asp Cys Ile Thr Arg Leu Ile His Leu Asn Lys
          50                  55                  60
Ile Asp Pro His Ala Pro Asn Glu Met Leu Tyr Gly Arg Ile Gly Tyr
      65                  70                  75
Ile Tyr Ala Leu Leu Phe Val Asn Lys Asn Phe Gly Val Glu Lys Thr
  80                  85                  90
Pro Gln Ser His Ile Gln Gln Ile Cys Glu Thr Ile Leu Thr Ser Gly
95                  100                 105                 110
Glu Asn Leu Ala Arg Lys Arg Asn Phe Thr Ala Lys Ser Pro Leu Met
              115                     120                     125
Tyr Glu Trp Tyr Gln Glu Tyr Tyr Val Gly Ala Ala His Gly Leu Ala
              130                     135                     140
Gly Ile Tyr Tyr Tyr Leu Met Gln Pro Ser Leu Gln Val Ser Gln Gly
          145                     150                     155
Lys Leu His Ser Leu Val Lys Pro Ser Val Asp Tyr Val Cys Gln Leu
      160                     165                     170
Lys Phe Pro Ser Gly Asn Tyr Pro Pro Cys Ile Gly Asp Asn Arg Asp
  175                     180                     185                     190
Leu Leu Val His Trp Cys His Gly Ala Pro Gly Val Ile Tyr Met Leu
                  195                     200                     205
Ile Gln Ala Tyr Lys Val Phe Arg Glu Glu Lys Tyr Leu Cys Asp Ala
              210                     215                     220
Tyr Gln Cys Ala Asp Val Ile Trp Gln Tyr Gly Leu Leu Lys Lys Gly
          225                     230                     235
```

```
Tyr Gly Leu Cys His Gly Ser Ala Gly Asn Ala Tyr Ala Phe Leu Thr
    240                 245                 250
Leu Tyr Asn Leu Thr Gln Asp Met Lys Tyr Leu Tyr Arg Ala Cys Lys
255                 260                 265                 270
Phe Ala Glu Trp Cys Leu Glu Tyr Gly Glu His Gly Cys Arg Thr Pro
                275                 280                 285
Asp Thr Pro Phe Ser Leu Phe Glu Gly Met Ala Gly Thr Ile Tyr Phe
                290                 295                 300
Leu Ala Asp Leu Leu Val Pro Thr Lys Ala Arg Phe Pro Ala Phe Glu
    305                 310                 315
Leu
```


```
<210> 197
<211> 30
<212> PRT
<213> Homo sapiens


<400> 197
Met Gln Met Asp Thr Phe Phe Met Ser Glu Lys His Thr His Thr His
1               5                   10                  15
Thr His Ile His Thr His Thr Arg Lys Thr Lys Lys Lys Lys
                20                  25                  30




<210> 198
<211> 112
<212> PRT
<213> Homo sapiens


<220>
<221> SIGNAL
<222> -48..-1


<400> 198
Met Gln Asp Thr Gly Ser Val Val Pro Leu His Trp Phe Gly Phe Gly
                        -45             -40                 -35
Tyr Ala Ala Leu Val Ala Ser Gly Gly Ile Ile Gly Tyr Val Lys Ala
        -30                 -25                 -20
Gly Ser Val Pro Ser Leu Ala Ala Gly Leu Leu Phe Gly Ser Leu Ala
        -15                 -10             .       -5
Gly Leu Gly Ala Tyr Gln Leu Ser Gln Asp Pro Arg Asn Val Trp Val
1               5                   10                  15
Phe Leu Ala Thr Ser Gly Thr Leu Ala Gly Ile Met Gly Met Arg Phe
                20                  25                  30
Tyr His Ser Gly Lys Phe Met Pro Ala Gly Leu Ile Ala Gly Ala Ser
        35                  40                  45
Leu Leu Met Val Ala Lys Val Gly Val Ser Met Phe Asn Arg Pro His
    50                  55                  60




<210> 199
<211> 54
<212> PRT
<213> Homo sapiens


<400> 199
Glu Ile Ala Gly Tyr Gly Ala Glu Gly Phe Ser Ser Val Leu Gly Tyr
1               5                   10                  15
```

```
Pro Arg Trp His Arg Leu Pro Pro Gln Ser Leu Gln His His Gln Tyr
            20                  25              30
Cys Gln Arg Arg Trp Pro Asp Arg Arg Cys Leu Gln Ser His Thr Gln
        35                  40              45
Ser Ser Gly His Leu Pro
    50
```


<210> 200
<211> 151
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 200
```
Met Ala Ala Ser Thr Ser Met Xaa Pro Val Ala Val Thr Ala Ala Val
    -20                  -15                  -10
Ala Pro Val Leu Ser Ile Asn Ser Asp Phe Ser Asp Leu Arg Glu Ile
-5                    1          5                    10
Lys Lys Gln Leu Leu Leu Ile Ala Gly Leu Thr Arg Glu Arg Gly Leu
            15                  20              25
Leu His Ser Ser Lys Trp Ser Ala Glu Leu Ala Phe Ser Leu Pro Ala
        30                  35              40
Leu Pro Xaa Gly Gln Leu Gln Pro Pro Pro Ile Thr Glu Glu Asp
    45                  50              55
Ala Gln Asp Met Asp Ala Tyr Thr Leu Ala Lys Ala Tyr Phe Asp Val
60                  65              70                  75
Lys Glu Tyr Asp Arg Ala Ala His Phe Leu His Gly Cys Asn Ser Lys
                80              85                  90
Lys Ala Tyr Phe Leu Tyr Met Tyr Ser Arg Tyr Leu Val Arg Ala Ile
            95                  100             105
Leu Lys Cys His Ser Ala Phe Ser Glu Thr Ser Ile Phe Arg Thr Asn
        110                 115             120
Gly Lys Val Lys Ser Phe Lys
    125             130
```


<210> 201
<211> 228
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -25..-1

<400> 201
```
Met Ser Met Ala Val Glu Thr Phe Gly Phe Phe Met Ala Thr Val Gly
-25                  -20                 -15                 -10
Leu Leu Met Leu Gly Val Thr Leu Pro Asn Ser Tyr Trp Arg Val Ser
            -5                  1               5
Thr Val His Gly Asn Val Ile Thr Thr Asn Thr Ile Phe Glu Asn Leu
        10                  15              20
Trp Phe Ser Cys Ala Thr Asp Ser Leu Gly Val Tyr Asn Cys Trp Glu
        25                  30              35
Phe Pro Ser Met Leu Ala Leu Ser Gly Tyr Ile Gln Ala Cys Arg Ala
40                  45              50                  55
```

```
Leu Met Ile Thr Ala Ile Leu Leu Gly Phe Leu Gly Leu Leu Leu Gly
            60                      65              70
Ile Ala Gly Leu Arg Cys Thr Asn Ile Gly Gly Leu Glu Leu Ser Arg
            75                  80                  85
Lys Ala Lys Leu Ala Ala Thr Ala Gly Ala Pro His Ile Leu Ala Gly
            90                  95                  100
Ile Cys Gly Met Val Ala Ile Ser Trp Tyr Ala Phe Asn Ile Thr Arg
        105             110                 115
Asp Phe Phe Asp Pro Leu Tyr Pro Gly Thr Lys Tyr Glu Leu Gly Pro
120                 125                 130                     135
Ala Leu Tyr Leu Gly Trp Ser Ala Ser Leu Ile Ser Ile Leu Gly Gly
                140                 145                 150
Leu Cys Leu Cys Ser Ala Cys Cys Cys Gly Ser Asp Glu Asp Pro Ala
                155                 160                 165
Ala Ser Ala Arg Arg Pro Tyr Gln Ala Pro Val Ser Val Met Pro Val
            170                 175                 180
Ala Thr Ser Asp Gln Glu Gly Asp Ser Ser Phe Gly Lys Tyr Gly Arg
            185                 190                     195
Asn Ala Tyr Val
200
```

```
<210> 202
<211> 64
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -47..-1

<400> 202
Met His Gly Phe Glu Ile Ile Ser Leu Lys Glu Glu Ser Pro Leu Gly
        -45                 -40                 -35
Lys Val Ser Gln Gly Pro Leu Phe Asn Val Thr Ser Gly Ser Ser Ser
        -30                 -25                 -20
Pro Val Thr Trp Leu Gly Leu Leu Ser Phe Gln Asn Leu His Cys Phe
-15                 -10                 -5                      1
Pro Asp Leu Pro Thr Glu Met Pro Leu Arg Ala Lys Gly Val Asn Thr
                5                   10                  15
```

```
<210> 203
<211> 146
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -31..-1

<400> 203
Met Met Trp Gln Lys Tyr Ala Gly Ser Arg Arg Ser Met Pro Leu Gly
        -30                 -25                 -20
Ala Arg Ile Leu Phe His Gly Val Phe Tyr Ala Gly Gly Phe Ala Ile
-15                 -10                 -5                      1
Val Tyr Tyr Leu Ile Gln Lys Phe His Ser Arg Ala Leu Tyr Tyr Lys
                5                   10                  15
Leu Ala Val Glu Gln Leu Gln Ser His Pro Glu Ala Gln Glu Ala Leu
            20                  25                  30
```

```
Gly Pro Pro Leu Asn Ile His Tyr Leu Lys Leu Ile Asp Arg Glu Asn
    35                  40                  45
Phe Val Asp Ile Val Asp Ala Lys Leu Lys Ile Pro Val Ser Gly Ser
50                  55                  60                  65
Lys Ser Glu Gly Leu Leu Tyr Val His Ser Ser Arg Gly Gly Pro Phe
                70                  75                  80
Gln Arg Trp His Leu Asp Glu Val Phe Leu Glu Leu Lys Asp Gly Gln
            85                  90                  95
Gln Ile Pro Val Phe Lys Leu Ser Gly Glu Asn Gly Asp Glu Val Lys
        100                 105                 110
Lys Glu
    115
```

```
<210> 204
<211> 87
<212> PRT
<213> Homo sapiens

<400> 204
Met Glu Leu Ala Pro Thr Ala Arg Leu Pro Pro Gly His Gly Ser Leu
1               5                   10                  15
Pro His Gly Val Leu Gly Pro Arg Ala Thr Gly Ser Val Thr His Leu
            20                  25                  30
Ser Leu Leu Pro Gln Ile Lys Gln Arg Ala Ser Glu Ala Leu Pro Glu
        35                  40                  45
Leu Leu Arg Pro Val Thr Pro Ile Thr Asn Phe Glu Gly Ser Gln Ser
    50                  55                  60
Gln Asp His Ser Gly Ile Phe Gly Leu Val Thr Asn Leu Glu Glu Leu
65                  70                  75                  80
Glu Val Asp Asp Trp Glu Phe
                85
```

```
<210> 205
<211> 40
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -27..-1

<400> 205
Met Arg Thr Leu Phe Gly Ala Val Arg Ala Pro Phe Ser Ser Leu Thr
        -25                 -20                 -15
Leu Leu Leu Ile Thr Pro Ser Pro Ser Pro Leu Leu Phe Asp Arg Gly
        -10                 -5                  1                   5
Leu Ser Leu Arg Ser Ala Met Ser
                10
```

```
<210> 206
<211> 154
<212> PRT
<213> Homo sapiens

<400> 206
Met Gly Ser Leu Ser Gly Leu Arg Leu Ala Ala Gly Ser Cys Phe Arg
```

```
1                5                          10                        15
Leu Cys Glu Arg Asp Val Ser Ser Ser Leu Arg Leu Thr Arg Ser Ser
            20                      25                  30
Asp Leu Lys Arg Ile Asn Gly Phe Cys Thr Lys Pro Gln Glu Ser Pro
            35              40                      45
Gly Ala Pro Ser Arg Thr Tyr Asn Arg Val Pro Leu His Lys Pro Thr
            50              55                      60
Asp Trp Gln Lys Lys Ile Leu Ile Trp Ser Gly Arg Phe Lys Lys Glu
65                      70                      75                      80
Asp Glu Ile Pro Glu Thr Val Ser Leu Glu Met Leu Asp Ala Ala Lys
                85                      90                      95
Asn Lys Met Arg Val Lys Ser Ser Tyr Leu Met Ile Ala Leu Thr Val
                100                     105                     110
Val Gly Cys Ile Phe Met Val Ile Glu Gly Lys Lys Ala Ala Gln Arg
            115                 120                     125
His Glu Thr Leu Thr Ser Leu Asn Leu Glu Lys Lys Ala Arg Leu Lys
        130                     135                     140
Glu Glu Ala Ala Met Lys Ala Lys Thr Glu
145                     150
```


```
<210> 207
<211> 101
<212> PRT
<213> Homo sapiens


<400> 207
Met Val Cys Glu Lys Cys Glu Lys Lys Leu Gly Thr Val Ile Thr Pro
1               5                       10                      15
Asp Thr Trp Lys Asp Gly Ala Arg Asn Thr Thr Glu Ser Gly Gly Arg
            20                      25                  30
Lys Leu Asn Lys Asn Lys Ala Leu Thr Ser Lys Lys Ala Arg Phe Asp
            35              40                      45
Pro Tyr Gly Lys Asn Lys Phe Ser Thr Cys Arg Ile Cys Lys Ser Ser
            50              55                      60
Val His Gln Pro Gly Ser His Tyr Cys Gln Gly Cys Ala Tyr Lys Lys
65                      70                      75                      80
Gly Ile Cys Ala Met Cys Gly Lys Lys Val Leu Asp Thr Lys Asn Tyr
                85                      90                      95
Lys Gln Thr Ser Val
            100
```


```
<210> 208
<211> 456
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -22..-1

<400> 208
Met Phe Glu Glu Pro Glu Trp Ala Glu Ala Ala Pro Val Ala Ala Gly
        -20                 -15                 -10
Leu Gly Pro Val Ile Ser Arg Pro Pro Pro Ala Ala Ser Ser Gln Asn
    -5                      1                   5                   10
Lys Gly Ser Lys Arg Arg Gln Leu Leu Ala Thr Leu Arg Ala Leu Glu
                15                  20                      25
Ala Ala Ser Leu Ser Gln His Pro Pro Ser Leu Cys Ile Ser Asp Ser
```

```
                 30                    35                    40
Glu Glu Glu Glu Glu Glu Arg Lys Lys Lys Cys Pro Lys Lys Ala Ser
         45                    50                    55
Phe Ala Ser Ala Ser Ala Glu Val Gly Lys Lys Gly Lys Lys Lys Cys
         60                    65                    70
Gln Lys Gln Gly Pro Pro Cys Ser Asp Ser Glu Glu Glu Val Glu Arg
75                    80                    85                    90
Lys Lys Lys Cys His Lys Gln Ala Leu Val Gly Ser Asp Ser Ala Glu
                 95                   100                   105
Asp Glu Lys Arg Lys Arg Lys Cys Gln Lys His Ala Pro Ile Asn Ser
             110                   115                   120
Ala Gln His Leu Asp Asn Val Asp Gln Thr Gly Pro Lys Ala Trp Lys
             125                   130                   135
Gly Ser Thr Thr Asn Asp Pro Pro Lys Gln Ser Pro Gly Ser Thr Ser
         140                     145                   150
Pro Lys Pro Pro His Thr Leu Ser Arg Lys Gln Trp Arg Asn Arg Gln
155                   160                   165                   170
Lys Asn Lys Arg Arg Cys Lys Asn Lys Phe Gln Pro Pro Gln Val Pro
                 175                   180                   185
Asp Gln Ala Pro Ala Glu Ala Pro Thr Glu Lys Thr Glu Val Ser Pro
             190                   195                   200
Val Pro Arg Thr Asp Ser His Gly Ala Arg Ala Gly Ala Leu Arg Ala
         205                   210                   215
Arg Met Ala Gln Arg Leu Asp Gly Ala Arg Phe Arg Tyr Leu Asn Glu
     220                   225                   230
Gln Leu Tyr Ser Gly Pro Ser Ser Ala Ala Gln Arg Leu Phe Gln Glu
235                   240                   245                   250
Asp Pro Glu Ala Phe Leu Leu Tyr His Arg Gly Phe Gln Ser Gln Val
             255                   260                   265
Lys Lys Trp Pro Leu Gln Pro Val Asp Arg Ile Ala Arg Asp Leu Arg
         270                   275                   280
Gln Arg Pro Ala Ser Leu Val Val Ala Asp Phe Gly Cys Gly Asp Cys
         285                   290                   295
Arg Leu Ala Ser Ser Ile Arg Asn Pro Val His Cys Phe Asp Leu Ala
     300                   305                   310
Ser Leu Asp Pro Arg Val Thr Val Cys Asp Met Ala Gln Val Pro Leu
315                   320                   325                   330
Glu Asp Glu Ser Val Asp Val Ala Val Phe Cys Leu Ser Leu Met Gly
             335                   340                   345
Thr Asn Ile Arg Asp Phe Leu Glu Glu Ala Asn Arg Val Leu Lys Pro
         350                   355                   360
Gly Gly Leu Leu Lys Val Ala Glu Val Ser Ser Arg Phe Glu Asp Val
         365                   370                   375
Arg Thr Phe Leu Arg Ala Val Thr Lys Leu Gly Phe Lys Ile Val Ser
     380                   385                   390
Lys Asp Leu Thr Asn Ser His Phe Phe Leu Phe Asp Phe Gln Lys Thr
395                   400                   405                   410
Gly Pro Pro Leu Val Gly Pro Lys Ala Gln Leu Ser Gly Leu Gln Leu
             415                   420                   425
Gln Pro Cys Leu Tyr Lys Arg Arg
             430
```

```
<210> 209
<211> 98
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -17..-1
```

<400> 209
Met Pro Ser Ser Phe Phe Leu Leu Leu Gln Phe Phe Leu Arg Ile Asp
        -15   -10   -5
Gly Val Leu Ile Arg Met Asn Asp Thr Arg Leu Tyr His Glu Ala Asp
    1   5   10   15
Lys Thr Tyr Met Leu Arg Glu Tyr Thr Ser Arg Glu Ser Lys Ile Ser
            20   25   30
Ser Leu Met His Val Pro Pro Ser Leu Phe Thr Glu Pro Asn Glu Ile
        35   40   45
Ser Gln Tyr Leu Pro Ile Lys Glu Ala Val Cys Glu Lys Leu Ile Phe
        50   55   60
Pro Glu Arg Ile Asp Pro Asn Pro Ala Asp Ser Gln Lys Ser Thr Gln
    65   70   75
Val Glu
80

<210> 210
<211> 83
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -29..-1

<400> 210
Met Thr Leu Leu Ser Phe Ala Ala Phe Thr Ala Ala Phe Ser Val Leu
            -25   -20   -15
Pro Cys Tyr Tyr Leu Gly Leu Phe Gln Arg Ala Leu Ala Ser Val Phe
        -10   -5   1
Asp Pro Leu Cys Val Cys Ser Arg Val Leu Pro Thr Pro Val Cys Thr
    5   10   15
Leu Val Ala Thr Gln Ala Glu Lys Ile Leu Glu Asn Gly Pro Cys Pro
20   25   30   35
Thr Lys Glu Ala Ala Gln Leu Val Gly Lys Gly Ser Val Ser Ala Arg
                40   45   50
Asn Ala Ser

<210> 211
<211> 229
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -23..-1

<400> 211
Met Gly Asp Lys Ile Trp Leu Pro Phe Pro Val Leu Leu Leu Ala Ala
            -20   -15   -10
Leu Pro Pro Val Leu Leu Pro Gly Ala Ala Gly Phe Thr Pro Ser Leu
        -5   1   5
Asp Ser Asp Phe Thr Phe Thr Leu Pro Ala Gly Gln Lys Glu Cys Phe
10   15   20   25
Tyr Gln Pro Met Pro Leu Lys Ala Ser Leu Glu Ile Glu Tyr Gln Val
            30   35   40
Leu Asp Gly Ala Gly Leu Asp Ile Asp Phe His Leu Ala Ser Pro Glu

```
                45                    50                    55
Gly Lys Thr Leu Val Phe Glu Gln Arg Lys Ser Asp Gly Val His Thr
        60                    65                    70
Val Glu Thr Glu Val Gly Asp Tyr Met Phe Cys Phe Asp Asn Thr Phe
    75                    80                    85
Ser Thr Ile Ser Glu Lys Val Ile Phe Phe Glu Leu Ile Leu Asp Asn
90                    95                    100                   105
Met Gly Glu Gln Ala Gln Glu Gln Glu Asp Trp Lys Lys Tyr Ile Thr
                110                   115                   120
Gly Thr Asp Ile Leu Asp Met Lys Leu Glu Asp Ile Leu Glu Ser Ile
            125                   130                   135
Ser Ser Ile Lys Ser Arg Leu Ser Lys Ser Gly His Ile Gln Ile Leu
        140                   145                   150
Leu Arg Ala Phe Glu Ala Arg Asp Arg Asn Ile Gln Glu Ser Asn Phe
    155                   160                   165
Asp Arg Val Asn Phe Trp Ser Met Val Asn Leu Val Val Met Val Val
170                   175                   180                   185
Val Ser Ala Ile Gln Val Tyr Met Leu Lys Ser Leu Phe Glu Asp Lys
                190                   195                   200
Arg Lys Ser Arg Thr
            205
```

<210> 212
<211> 152
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

```
<400> 212
Met Ala Gln Leu Gly Ala Val Val Ala Val Ala Ser Ser Phe Phe Cys
        -20                   -15                   -10
Ala Ser Leu Phe Ser Ala Val His Lys Ile Glu Glu Gly His Ile Gly
-5                    1                     5                     10
Val Tyr Tyr Arg Gly Gly Ala Leu Leu Thr Ser Thr Ser Gly Pro Gly
                15                    20                    25
Phe His Leu Met Leu Pro Phe Ile Thr Ser Tyr Lys Ser Val Gln Thr
            30                    35                    40
Thr Leu Gln Thr Asp Glu Val Lys Asn Val Pro Cys Gly Thr Ser Gly
    45                    50                    55
Gly Val Met Ile Tyr Phe Asp Arg Ile Glu Val Val Asn Phe Leu Val
60                    65                    70                    75
Pro Asn Ala Val His Asp Ile Val Lys Asn Tyr Thr Ala Asp Tyr Asp
                80                    85                    90
Lys Ala Leu Ile Phe Asn Lys Ile His His Glu Leu Asn Gln Phe Cys
            95                    100                   105
Ser Val His Thr Leu Gln Glu Val Tyr Ile Glu Leu Phe Gly Leu Glu
        110                   115                   120
Asn Asp Phe Ser Gln Glu Ser Ser
    125                   130
```

<210> 213
<211> 179
<212> PRT
<213> Homo sapiens

```
<220>
<221> SIGNAL
<222> -54..-1

<400> 213
Met Ala Ala Ser Glu Ala Ala Val Val Ser Ser Pro Ser Leu Lys Thr
              -50                   -45                   -40
Asp Thr Ser Pro Val Leu Glu Thr Ala Gly Thr Val Ala Ala Met Ala
              -35                   -30                   -25
Ala Thr Pro Ser Ala Arg Ala Ala Ala Ala Val Val Ala Ala Ala Ala
          -20                   -15                   -10
Arg Thr Gly Ser Glu Ala Arg Val Ser Lys Ala Ala Leu Ala Thr Lys
     -5              1                 5                     10
Leu Leu Ser Leu Ser Gly Val Phe Ala Val His Lys Pro Lys Gly Pro
              15                  20                    25
Thr Ser Ala Glu Leu Leu Asn Arg Leu Lys Glu Lys Leu Leu Ala Glu
              30                  35                    40
Ala Gly Met Pro Ser Pro Glu Trp Thr Lys Arg Lys Lys Gln Thr Leu
          45                  50                    55
Lys Ile Gly His Gly Gly Thr Leu Asp Ser Ala Ala Arg Gly Val Leu
     60              65                  70
Val Val Gly Ile Gly Ser Gly Thr Lys Met Leu Thr Ser Met Leu Ser
75              80                  85                    90
Gly Ser Lys Arg Tyr Thr Ala Ile Gly Glu Leu Gly Lys Ala Thr Asp
              95                  100                   105
Thr Leu Asp Ser Thr Gly Lys Val Thr Glu Glu Lys Pro Tyr Gly Met
          110                 115                   120
Asn Leu Ile
          125


<210> 214
<211> 269
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -92..-1

<400> 214
Met Ile Thr His Val Thr Leu Glu Asp Ala Leu Ser Asn Val Asp Leu
          -90                 -85                   -80
Leu Glu Glu Leu Pro Leu Pro Asp Gln Gln Pro Cys Ile Glu Pro Pro
     -75                 -70                   -65
Pro Ser Ser Ile Met Tyr Gln Ala Asn Phe Asp Thr Asn Phe Glu Asp
-60                 -55                   -50                   -45
Arg Asn Ala Phe Val Thr Gly Ile Ala Arg Tyr Ile Glu Gln Ala Thr
              -40                 -35                   -30
Val His Ser Ser Met Asn Glu Met Leu Glu Glu Gly His Glu Tyr Ala
          -25                 -20                   -15
Val Met Leu Tyr Thr Trp Arg Ser Cys Ser Arg Ala Ile Pro Gln Val
          -10                 -5                  1
Lys Cys Asn Glu Gln Pro Asn Arg Val Glu Ile Tyr Glu Lys Thr Val
5                   10                  15                    20
Glu Val Leu Glu Pro Glu Val Thr Lys Leu Met Lys Phe Met Tyr Phe
              25                  30                    35
Gln Arg Lys Ala Ile Glu Arg Phe Cys Ser Glu Val Lys Arg Leu Cys
              40                  45                    50
His Ala Glu Arg Arg Lys Asp Phe Val Ser Glu Ala Tyr Leu Leu Thr
          55                  60                    65
```

```
Leu Gly Lys Phe Ile Asn Met Phe Ala Val Leu Asp Glu Leu Lys Asn
    70              75              80
Met Lys Cys Ser Val Lys Asn Asp His Ser Ala Tyr Lys Arg Ala Ala
85              90              95              100
Gln Phe Leu Arg Lys Met Ala Asp Pro Gln Ser Ile Gln Glu Ser Gln
            105             110             115
Asn Leu Ser Met Phe Leu Ala Asn His Asn Arg Ile Thr Gln Cys Leu
            120             125             130
His Gln Gln Leu Glu Val Ile Pro Gly Tyr Glu Glu Leu Leu Ala Asp
            135             140             145
Ile Val Asn Ile Cys Val Asp Tyr Tyr Glu Asn Lys Met Tyr Leu Thr
    150             155             160
Pro Ser Glu Lys His Met Leu Leu Lys Val Lys Leu Pro
165             170             175
```

```
<210> 215
<211> 135
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -22..-1

<400> 215
Met Gln Thr Val Tyr Tyr Gly Ser Leu Gly Leu Trp Leu Ala Leu Val
        -20             -15             -10
Asp Gly Leu Val Arg Ser Ser Pro Ser Leu Asp Gln Met Phe Asp Ala
    -5              1               5               10
Glu Ile Leu Gly Phe Ser Thr Pro Pro Gly Arg Leu Ser Met Met Ser
            15              20              25
Phe Ile Phe Asn Ala Leu Thr Cys Ala Leu Gly Leu Leu Tyr Phe Ile
            30              35              40
Arg Arg Gly Lys Gln Cys Leu Asp Phe Thr Val Thr Val His Phe Phe
            45              50              55
His Leu Leu Gly Cys Trp Phe Tyr Ser Ser Arg Phe Pro Ser Ala Leu
    60              65              70
Thr Trp Trp Leu Val Gln Ala Val Cys Ile Ala Leu Met Ala Val Ile
75              80              85              90
Gly Glu Tyr Leu Cys Met Arg Thr Glu Leu Lys Glu Ile Pro Leu Asn
            95              100             105
Ser Ala Pro Lys Ser Asn Val
            110
```

```
<210> 216
<211> 67
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -38..-1

<400> 216
Met Asn Asn Val Gln Pro Lys Ile Lys His Arg Pro Phe Cys Phe Ser
            -35             -30             -25
Val Lys Gly His Val Lys Met Leu Arg Leu Val Phe Ala Leu Val Thr
        -20             -15             -10
```

```
Ala Val Cys Cys Leu Ala Asp Gly Ala Leu Ile Tyr Arg Lys Leu Leu
     -5                   1             5                       10
Phe Asn Pro Asn Gly Pro Tyr Gln Lys Lys Pro Val His Glu Lys Lys
             15                  20                  25
Glu Val Leu
```

```
<210> 217
<211> 125
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -54..-1

<400> 217
Met Ala Asp Glu Glu Leu Glu Ala Leu Arg Arg Gln Arg Leu Ala Glu
             -50                 -45                 -40
Leu Gln Ala Lys His Gly Asp Pro Gly Asp Ala Ala Gln Gln Glu Ala
             -35                 -30                 -25
Lys His Arg Glu Ala Glu Met Arg Asn Ser Ile Leu Ala Gln Val Leu
         -20                 -15                 -10
Asp Gln Ser Ala Arg Ala Arg Leu Ser Asn Leu Ala Leu Val Lys Pro
     -5                   1             5                       10
Glu Lys Thr Lys Ala Val Glu Asn Tyr Leu Ile Gln Met Ala Arg Tyr
             15                  20                  25
Gly Gln Leu Ser Glu Lys Val Ser Glu Gln Gly Leu Ile Glu Ile Leu
             30                  35                  40
Lys Lys Val Ser Gln Gln Thr Glu Lys Thr Thr Thr Val Lys Phe Asn
         45                  50                  55
Arg Arg Lys Val Met Asp Ser Asp Glu Asp Asp Asp Tyr
     60                  65                  70
```

```
<210> 218
<211> 376
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 218
Met Gly His Arg Phe Leu Arg Gly Leu Leu Thr Leu Leu Leu Pro Pro
         -20                 -15                 -10
Pro Pro Leu Tyr Thr Arg His Arg Met Leu Gly Pro Glu Ser Val Pro
-5                   1             5                       10
Pro Pro Lys Arg Ser Arg Ser Lys Leu Met Ala Pro Pro Arg Ile Gly
             15                  20                  25
Thr His Asn Gly Thr Phe His Cys Asp Glu Ala Leu Ala Cys Ala Leu
         30                  35                  40
Leu Arg Leu Leu Pro Glu Tyr Arg Asp Ala Glu Ile Val Arg Thr Arg
         45                  50                  55
Asp Pro Glu Lys Leu Ala Ser Cys Asp Ile Val Val Asp Val Gly Gly
60                  65                  70                      75
Glu Tyr Asp Pro Arg Arg His Arg Tyr Asp His His Gln Arg Ser Phe
             80                  85                  90
Thr Glu Thr Met Ser Ser Leu Ser Pro Gly Arg Pro Trp Gln Thr Lys
```

```
                95                      100                    105
Leu Ser Ser Ala Gly Leu Ile Tyr Leu His Phe Gly His Lys Leu Leu
        110                    115                    120
Ala Gln Leu Leu Gly Thr Ser Glu Glu Asp Ser Met Val Gly Thr Leu
    125                    130                    135
Tyr Asp Lys Met Tyr Glu Asn Phe Val Glu Glu Val Asp Ala Val Asp
140                    145                    150                    155
Asn Gly Ile Ser Gln Trp Ala Glu Gly Glu Pro Arg Tyr Ala Leu Thr
            160                    165                    170
Thr Thr Leu Ser Ala Arg Val Ala Arg Leu Asn Pro Thr Trp Asn His
        175                    180                    185
Pro Asp Gln Asp Thr Glu Ala Gly Phe Lys Arg Ala Met Asp Leu Val
        190                    195                    200
Gln Glu Glu Phe Leu Gln Arg Leu Asp Phe Tyr Gln His Ser Trp Leu
    205                    210                    215
Pro Ala Arg Ala Leu Val Glu Glu Ala Leu Ala Gln Arg Phe Gln Val
220                    225                    230                    235
Asp Pro Ser Gly Glu Ile Val Glu Leu Ala Lys Gly Ala Cys Pro Trp
            240                    245                    250
Lys Glu His Leu Tyr His Leu Glu Ser Gly Leu Ser Pro Pro Val Ala
        255                    260                    265
Ile Phe Phe Val Ile Tyr Thr Asp Gln Ala Gly Gln Trp Arg Ile Gln
        270                    275                    280
Cys Val Pro Lys Glu Pro His Ser Phe Gln Ser Arg Leu Pro Leu Pro
    285                    290                    295
Glu Pro Trp Arg Gly Leu Arg Asp Glu Ala Leu Asp Gln Val Ser Gly
300                    305                    310                    315
Ile Pro Gly Cys Ile Phe Val His Ala Ser Gly Phe Ile Gly Gly His
            320                    325                    330
Arg Thr Arg Glu Gly Ala Leu Ser Met Ala Arg Ala Thr Leu Ala Gln
        335                    340                    345
Arg Ser Tyr Leu Pro Gln Ile Ser
        350                    355
```

```
<210> 219
<211> 211
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -30..-1

<400> 219
Met Gly Glu Ala Ser Pro Pro Ala Pro Ala Arg Arg His Leu Leu Val
-30                    -25                    -20                    -15
Leu Leu Leu Leu Leu Ser Thr Leu Val Ile Pro Ser Ala Ala Ala Pro
            -10                    -5                     1
Ile His Asp Ala Asp Ala Gln Glu Ser Ser Leu Gly Leu Thr Gly Leu
        5                      10                     15
Gln Ser Leu Leu Gln Gly Phe Ser Arg Leu Phe Leu Lys Gly Asn Leu
    20                     25                     30
Leu Arg Gly Ile Asp Ser Leu Phe Ser Ala Pro Met Asp Phe Arg Gly
35                     40                     45                     50
Leu Pro Gly Asn Tyr His Lys Glu Glu Asn Gln Glu His Gln Leu Gly
            55                     60                     65
Asn Asn Thr Leu Ser Ser His Leu Gln Ile Asp Lys Val Pro Arg Met
            70                     75                     80
Glu Glu Lys Glu Ala Leu Val Pro Ile Gln Lys Ala Thr Asp Ser Phe
        85                     90                     95
```

```
His Thr Glu Leu His Pro Arg Val Ala Phe Trp Ile Ile Lys Leu Pro
    100             105             110
Arg Arg Arg Ser His Gln Asp Ala Leu Glu Gly Gly His Trp Leu Ser
115             120             125             130
Glu Lys Arg His Arg Leu Gln Ala Ile Arg Asp Gly Leu Arg Lys Gly
            135             140             145
Thr His Lys Asp Val Leu Glu Glu Gly Thr Glu Ser Ser Ser His Ser
            150             155             160
Arg Leu Ser Pro Arg Lys Thr His Leu Leu Tyr Ile Leu Arg Pro Ser
        165             170             175
Arg Gln Leu
    180
```

<210> 220
<211> 154
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -60..-1

<400> 220
```
Met Gly Ser Lys Cys Cys Lys Gly Gly Pro Asp Glu Asp Ala Val Glu
-60             -55             -50             -45
Arg Gln Arg Arg Gln Lys Leu Leu Leu Ala Gln Leu His His Arg Lys
            -40             -35             -30
Arg Val Lys Ala Ala Gly Gln Ile Gln Ala Trp Trp Arg Gly Val Leu
            -25             -20             -15
Val Arg Arg Thr Leu Leu Val Ala Ala Leu Arg Ala Trp Met Ile Gln
            -10             -5              1
Cys Trp Trp Arg Thr Leu Val Gln Arg Arg Ile Arg Gln Arg Arg Gln
5               10              15              20
Ala Leu Leu Arg Val Tyr Val Ile Gln Glu Gln Ala Thr Val Lys Leu
            25              30              35
Gln Ser Cys Ile Arg Met Trp Gln Cys Arg Gln Cys Tyr Arg Gln Met
            40              45              50
Cys Asn Ala Leu Cys Leu Phe Gln Val Pro Glu Ser Ser Leu Ala Phe
            55              60              65
Gln Thr Asp Gly Phe Leu Gln Val Gln Tyr Ala Ile Pro Ser Lys Gln
    70              75              80
Pro Glu Phe His Ile Glu Ile Leu Ser Ile
85              90
```

<210> 221
<211> 123
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -42..-1

<400> 221
```
Met Lys Gly Gly Ala Phe Ser Asn Leu Asn Asp Ser Gln Leu Ser Ala
        -40             -35             -30
Ser Phe Leu Gln Pro Ser Leu Gln Ala Asn Cys Pro Ala Leu Asp Pro
    -25             -20             -15
```

```
Ala Val Ser Leu Ser Ala Pro Ala Phe Ala Ser Ala Leu Arg Ser Met
-10                 -5                 1                 5
Lys Ser Ser Gln Ala Ala Arg Lys Asp Asp Phe Leu Arg Ser Leu Ser
            10              15              20
Asp Gly Asp Ser Gly Thr Ser Glu His Ile Ser Ala Val Val Thr Ser
        25              30              35
Pro Arg Ile Ser Cys His Gly Ala Ala Ile Pro Thr Ala Arg Ala Leu
    40              45              50
Cys Leu Gly Cys Ser Cys Cys Thr Glu Arg Leu Leu Leu Pro Pro Pro
55              60              65              70
Ser Leu Leu Ser Leu Glu Ala Pro Ala Ser Thr
            75              80
```

```
<210> 222
<211> 346
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -19..-1

<400> 222
Met Ala Met Ala Gln Lys Leu Ser His Leu Leu Pro Ser Leu Arg Gln
            -15             -10             -5
Val Ile Gln Glu Pro Gln Leu Ser Leu Gln Pro Glu Pro Val Phe Thr
            1               5               10
Val Asp Arg Ala Glu Val Pro Pro Leu Phe Trp Lys Pro Tyr Ile Tyr
    15              20              25
Ala Gly Tyr Arg Pro Leu His Gln Thr Trp Arg Phe Tyr Phe Arg Thr
30              35              40              45
Leu Phe Gln Gln His Asn Glu Ala Val Asn Val Trp Thr His Leu Leu
            50              55              60
Ala Ala Leu Val Leu Leu Leu Arg Leu Ala Leu Phe Val Glu Thr Val
            65              70              75
Asp Phe Trp Gly Asp Pro His Ala Leu Pro Leu Phe Ile Ile Val Leu
        80              85              90
Ala Ser Phe Thr Tyr Leu Ser Leu Ser Ala Leu Ala His Leu Leu Gln
    95              100             105
Ala Lys Ser Glu Phe Trp His Tyr Ser Phe Phe Phe Leu Asp Tyr Val
110             115             120             125
Gly Val Ala Val Tyr Gln Phe Gly Ser Ala Leu Ala His Phe Tyr Tyr
            130             135             140
Ala Ile Glu Pro Ala Trp His Ala Gln Val Gln Ala Val Phe Leu Pro
            145             150             155
Met Ala Ala Phe Leu Ala Trp Leu Ser Cys Ile Gly Ser Cys Tyr Asn
        160             165             170
Lys Tyr Ile Gln Lys Pro Gly Leu Leu Gly Arg Thr Cys Gln Glu Val
    175             180             185
Pro Ser Val Leu Ala Tyr Ala Leu Asp Ile Ser Pro Val Val His Arg
190             195             200             205
Ile Phe Val Ser Ser Asp Pro Thr Thr Asp Asp Pro Ala Leu Leu Tyr
            210             215             220
His Lys Cys Gln Val Val Phe Phe Leu Leu Ala Ala Ala Phe Phe Ser
        225             230             235
Thr Phe Met Pro Glu Arg Trp Phe Pro Gly Ser Cys His Val Phe Gly
        240             245             250
Gln Gly His Gln Leu Phe His Ile Phe Leu Val Leu Cys Thr Leu Ala
    255             260             265
Gln Leu Glu Ala Val Ala Leu Asp Tyr Glu Ala Arg Arg Pro Ile Tyr
```

```
270                          275                          280                          285
Glu Pro Leu His Thr His Trp Pro His Asn Phe Ser Gly Leu Phe Leu
                    290                          295                          300
Leu Thr Val Gly Ser Ser Ile Leu Thr Ala Phe Leu Leu Ser Gln Leu
                    305                          310                          315
Val Gln Arg Lys Leu Asp Gln Lys Thr Lys
                    320                          325
```

<210> 223
<211> 210
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -20..-1

<400> 223
```
Met Asp Asn Arg Phe Ala Thr Ala Phe Val Ile Ala Cys Val Leu Ser
-20                          -15                          -10                          -5
Leu Ile Ser Thr Ile Tyr Met Ala Ala Ser Ile Gly Thr Asp Phe Trp
                    1                            5                            10
Tyr Glu Tyr Arg Ser Pro Val Gln Glu Asn Ser Ser Asp Leu Asn Lys
        15                           20                           25
Ser Ile Trp Asp Glu Phe Ile Ser Asp Glu Ala Asp Glu Lys Thr Tyr
        30                           35                           40
Asn Asp Ala Leu Phe Arg Tyr Asn Gly Thr Val Gly Leu Trp Arg Arg
45                           50                           55                           60
Cys Ile Thr Ile Pro Lys Asn Met His Trp Tyr Ser Pro Pro Glu Arg
                    65                           70                           75
Thr Glu Ser Phe Asp Val Val Thr Lys Cys Val Ser Phe Thr Leu Thr
                    80                           85                           90
Glu Gln Phe Met Glu Lys Phe Val Asp Pro Gly Asn His Asn Ser Gly
        95                           100                          105
Ile Asp Leu Leu Arg Thr Tyr Leu Trp Arg Cys Gln Phe Leu Leu Pro
        110                          115                          120
Phe Val Ser Leu Gly Leu Met Cys Phe Gly Ala Leu Ile Gly Leu Cys
125                          130                          135                          140
Ala Cys Ile Cys Arg Ser Leu Tyr Pro Thr Ile Ala Thr Gly Ile Leu
                    145                          150                          155
His Leu Leu Ala Val Thr Lys Glu Ser Met Leu Pro Ala Gly Ala Glu
                    160                          165                          170
Ser Lys His Thr Ala Thr Pro Ala His Ala Cys Val Gln Thr Gly Lys
        175                          180                          185
Pro Lys
    190
```

<210> 224
<211> 184
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -20..-1

<400> 224
```
Met Asp Asn Arg Phe Ala Thr Ala Phe Val Ile Ala Cys Val Leu Ser
```

```
-20                  -15                          -10                          -5
Leu Ile Ser Thr Ile Tyr Met Ala Ala Ser Ile Gly Thr Asp Phe Trp
              1                   5                        10
Tyr Glu Tyr Arg Ser Pro Val Gln Glu Asn Ser Ser Asp Leu Asn Lys
          15                  20                      25
Ser Ile Trp Asp Glu Phe Ile Ser Asp Glu Ala Asp Glu Lys Thr Tyr
    30                  35                      40
Asn Asp Ala Pro Phe Arg Tyr Asn Gly Thr Val Gly Leu Trp Arg Arg
45                  50                      55                      60
Cys Ile Thr Ile Pro Lys Asn Met His Trp Tyr Ser Pro Pro Glu Arg
              65                  70                      75
Thr Glu Ser Phe Asp Val Val Thr Lys Cys Val Ser Phe Thr Leu Thr
              80                  85                      90
Glu Gln Phe Met Glu Lys Phe Val Asp Pro Gly Asn His Asn Ser Gly
          95                  100 .                   105
Ile Asp Leu Leu Arg Thr Tyr Leu Trp Arg Cys Gln Phe Leu Leu Pro
    110                 115                     120
Phe Val Ser Leu Gly Leu Met Cys Phe Gly Ala Leu Ile Gly Leu Cys
125                 130                     135                     140
Ala Cys Ile Cys Arg Ser Leu Tyr Pro Thr Ile Ala Thr Gly Ile Leu
              145                 150                     155
His Leu Leu Ala Asp Thr Met Leu
              160
```

<210> 225
<211> 227
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -22..-1

<400> 225

```
Met Gly Trp Thr Met Arg Leu Val Thr Ala Ala Leu Leu Leu Gly Leu
          -20                 -15                     -10
Met Met Val Val Thr Gly Asp Glu Asp Glu Asn Ser Pro Cys Ala His
    -5                  1                   5                       10
Glu Ala Leu Leu Asp Glu Asp Thr Leu Phe Cys Gln Gly Leu Glu Val
              15                  20                      25
Phe Tyr Pro Glu Leu Gly Asn Ile Gly Cys Lys Val Val Pro Asp Cys
          30                  35                      40
Asn Asn Tyr Arg Gln Lys Ile Thr Ser Trp Met Glu Pro Ile Val Lys
          45                  50                  .   55
Phe Pro Gly Ala Val Asp Gly Ala Thr Tyr Ile Leu Val Met Val Asp
          60                  65                      70
Pro Asp Ala Pro Ser Arg Ala Glu Pro Arg Gln Arg Phe Trp Arg His
75                  80                      85                      90
Trp Leu Val Thr Asp Ile Lys Gly Ala Asp Leu Lys Lys Gly Lys Ile
              95                  100                     105
Gln Gly Gln Glu Leu Ser Ala Tyr Gln Ala Pro Ser Pro Pro Ala His
          110                 115                     120
Ser Gly Phe His Arg Tyr Gln Phe Phe Val Tyr Leu Gln Glu Gly Lys
          125                 130                     135
Val Ile Ser Leu Leu Pro Lys Glu Asn Lys Thr Arg Gly Ser Trp Lys
    140                 145                     150
Met Asp Arg Phe Leu Asn Arg Phe His Leu Gly Glu Pro Glu Ala Ser
155                 160                     165                     170
Thr Gln Phe Met Thr Gln Asn Tyr Gln Asp Ser Pro Thr Leu Gln Ala
              175                 180                     185
```

Pro Arg Glu Arg Ala Ser Glu Pro Lys His Lys Asn Gln Ala Glu Ile
            190   195   200
Ala Ala Cys
        205

<210> 226
<211> 74
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -41..-1

<400> 226
Met Ile Ala Arg Arg Asn Pro Val Pro Leu Arg Phe Leu Pro Asp Glu
    -40   -35   -30
Ala Arg Ser Leu Pro Pro Pro Lys Leu Thr Asp Pro Arg Leu Leu Tyr
-25   -20   -15   -10
Ile Gly Phe Leu Gly Tyr Cys Ser Gly Leu Ile Asp Asn Leu Ile Arg
                -5   1   5
Arg Arg Pro Ile Ala Thr Ala Gly Leu His Arg Gln Leu Leu Tyr Ile
        10   15   20
Thr Ala Phe Phe Leu Leu Asp Ile Ile Leu
    25   30

<210> 227
<211> 73
<212> PRT
<213> Homo sapiens

<400> 227
Met Glu Lys Tyr Glu Asn Leu Gly Leu Val Gly Glu Gly Ser Tyr Gly
1   5   10   15
Met Val Met Lys Cys Arg Asn Lys Asp Thr Gly Arg Ile Val Ala Ile
            20   25   30
Lys Lys Phe Leu Glu Ser Asp Asp Asp Lys Met Val Lys Lys Ile Ala
        35   40   45
Met Arg Glu Val Lys Leu Leu Lys Gln Leu Arg His Glu Asn Leu Val
    50   55   60
Asn Leu Leu Glu Val Cys Lys Lys Lys
65   70

<210> 228
<211> 82
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -16..-1

<400> 228
Met Lys Arg Leu Leu Pro Ala Thr Ser Leu Ala Gly Pro Val Leu Ser
    -15   -10   -5
Thr Leu Ile Ala Pro Thr Pro Met Leu Phe Cys Glu Asp Lys Ser Trp

```
1                 5                      10                     15
Asp Leu Phe Leu Phe Phe Lys Ser His Lys Thr Trp Gly Ile Ser Thr
              20                  25                  30
Asn Leu Ser Ser Cys Pro Phe Gly Asn Leu Phe Leu Cys Val Gln Phe
           35                  40                  45
Val Arg Glu Lys Gln Ser Phe Cys Met Asn Thr Glu Cys Asp Leu Arg
        50                  55                  60
Lys Asn
65
```

```
<210> 229
<211> 119
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -56..-1

<400> 229
Met Ala Glu Pro Ser Ala Ala Thr Gln Ser His Ser Ile Ser Ser Ser
    -55                 -50                 -45
Ser Phe Gly Ala Glu Pro Ser Ala Pro Gly Gly Gly Gly Ser Pro Gly
-40                 -35                 -30                 -25
Ala Cys Pro Ala Leu Gly Thr Lys Ser Cys Ser Ser Ser Cys Ala Asp
                -20                 -15                 -10
Ser Phe Val Ser Ser Ser Ser Ser Gln Pro Val Ser Leu Phe Ser Thr
                -5                  1                   5
Ser Gln Glu Gly Leu Ser Ser Leu Cys Ser Asp Glu Pro Ser Ser Glu
        10                  15                  20
Ile Met Thr Ser Ser Phe Leu Ser Ser Ser Glu Ile His Asn Thr Gly
25                  30                  35                  40
Leu Thr Ile Leu His Gly Glu Lys Ser His Val Leu Gly Ser Gln Pro
                45                  50                  55
Ile Leu Ala Lys Lys Lys Lys
            60
```

```
<210> 230
<211> 54
<212> PRT
<213> Homo sapiens

<400> 230
Ala Phe Val Trp Glu Pro Ala Met Val Arg Ile Asn Ala Leu Thr Ala
1               5                   10                  15
Ala Ser Glu Ala Ala Cys Leu Ile Val Ser Val Asp Glu Thr Ile Lys
            20                  25                  30
Asn Pro Arg Ser Thr Val Asp Ala Pro Thr Ala Ala Gly Arg Gly Arg
        35                  40                  45
Gly Arg Gly Arg Pro His
    50
```

```
<210> 231
<211> 210
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -14..-1

<400> 231
Met Leu Thr Leu Leu Gly Leu Ser Phe Ile Leu Ala Gly Leu Ile Val
            -10              -5                   1
Gly Gly Ala Cys Ile Tyr Lys Tyr Phe Met Pro Lys Ser Thr Ile Tyr
         5              10              15
Arg Gly Glu Met Cys Phe Phe Asp Ser Glu Asp Pro Ala Asn Ser Leu
      20              25              30
Arg Gly Gly Glu Pro Asn Phe Leu Pro Val Thr Glu Glu Ala Asp Ile
35              40              45              50
Arg Glu Asp Asp Asn Ile Ala Ile Ile Asp Val Pro Val Pro Ser Phe
               55              60              65
Ser Asp Ser Asp Pro Ala Ala Ile Ile His Asp Phe Glu Lys Gly Met
            70              75              80
Thr Ala Tyr Leu Asp Leu Leu Leu Gly Ile Cys Tyr Leu Met Pro Leu
         85              90              95
Asn Thr Ser Ile Val Met Pro Pro Lys Asn Leu Val Glu Leu Phe Gly
      100             105             110
Lys Leu Ala Ser Gly Arg Tyr Leu Pro Gln Thr Tyr Val Val Arg Glu
115             120             125             130
Asp Leu Val Ala Val Glu Glu Ile Arg Asp Val Ser Asn Leu Gly Ile
               135             140             145
Phe Ile Tyr Gln Leu Cys Asn Asn Arg Lys Ser Phe Arg Leu Arg Arg
            150             155             160
Arg Asp Leu Leu Leu Gly Phe Asn Lys Arg Ala Ile Asp Lys Cys Trp
         165             170             175
Lys Ile Arg His Phe Pro Asn Glu Phe Ile Val Glu Thr Lys Ile Cys
         180             185             190
Gln Glu
195



<210> 232
<211> 108
<212> PRT
<213> Homo sapiens

<400> 232
Met Gly Cys Val Phe Gln Ser Thr Glu Asp Lys Cys Ile Phe Lys Ile
1              5              10              15
Asp Trp Thr Leu Ser Pro Gly Glu His Ala Lys Asp Glu Tyr Val Leu
         20              25              30
Tyr Tyr Tyr Ser Asn Leu Ser Val Pro Ile Gly Arg Phe Gln Asn Arg
         35              40              45
Val His Leu Met Gly Asp Ile Leu Cys Asn Asp Gly Ser Leu Leu Leu
      50              55              60
Gln Asp Val Gln Glu Ala Asp Gln Gly Thr Tyr Ile Cys Glu Ile Arg
65              70              75              80
Leu Lys Gly Glu Ser Gln Val Phe Lys Lys Ala Val Val Leu His Val
               85              90              95
Leu Pro Glu Glu Pro Lys Gly Thr Gln Met Leu Thr
            100             105



<210> 233
<211> 43
```

<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -18...-1

<400> 233
Met Ser Ser Gly Arg Leu Arg Trp Leu Met Pro Val Ile Pro Ala Leu
          -15   -10   -5
Trp Gly Ala Glu Lys Gly Glu Ser Pro Glu Val Ser Ser Phe Glu Thr
          1   5   10
Arg Leu Ala Asn Met Ala Lys Pro Cys Leu Tyr
15   20   25

<210> 234
<211> 36
<212> PRT
<213> Homo sapiens

<400> 234
Met Ser Ala Arg Ile Pro Phe Tyr Lys Asp Thr Ser Gln Ile Arg Leu
1   5   10   15
Gly Ser Thr Ile Ile Pro His Phe Asn Leu Ile Thr Phe Val Lys Thr
          20   25   30
Phe Phe Gln Ile
          35

<210> 235
<211> 307
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -13...-1

<400> 235
Met Leu Ala Val Ser Leu Thr Val Pro Leu Leu Gly Ala Met Met Leu
          -10   -5   1
Leu Glu Ser Pro Ile Asp Pro Gln Pro Leu Ser Phe Lys Glu Pro Pro
    5   10   15
Leu Leu Leu Gly Val Leu His Pro Asn Thr Lys Leu Arg Gln Ala Glu
20   25   30   35
Arg Leu Phe Glu Asn Gln Leu Val Gly Pro Glu Ser Ile Ala His Ile
          40   45   50
Gly Asp Val Met Phe Thr Gly Thr Ala Asp Gly Arg Val Val Lys Leu
          55   60   65
Glu Asn Gly Glu Ile Glu Thr Ile Ala Arg Phe Gly Ser Gly Pro Cys
          70   75   80
Lys Thr Arg Asp Asp Glu Pro Val Cys Gly Arg Pro Leu Gly Ile Arg
    85   90   95
Ala Gly Pro Asn Gly Thr Leu Phe Val Ala Asp Ala Cys Lys Gly Leu
100   105   110   115
Phe Glu Val Asn Pro Trp Lys Arg Glu Val Lys Leu Leu Leu Ser Ser
          120   125   130
Glu Thr Pro Ile Glu Gly Lys Asn Met Ser Phe Val Asn Asp Leu Thr
          135   140   145

```
Val Ser Gln Asp Gly Arg Lys Ile Tyr Phe Thr Asp Ser Ser Ser Lys
        150                 155                 160
Trp Gln Arg Arg Asp Tyr Leu Leu Leu Val Met Glu Gly Thr Asp Asp
        165                 170                 175
Gly Arg Leu Leu Glu Tyr Asp Thr Val Thr Arg Glu Val Lys Val Leu
180                 185                 190                 195
Leu Asp Gln Leu Arg Phe Pro Asn Gly Val Gln Leu Ser Pro Ala Glu
                200                 205                 210
Asp Phe Val Leu Val Ala Glu Thr Thr Met Ala Arg Ile Arg Arg Val
                215                 220                 225
Tyr Val Ser Gly Leu Met Lys Gly Gly Ala Asp Leu Phe Val Glu Asn
        230                 235                 240
Met Pro Gly Phe Pro Asp Asn Ile Arg Pro Ser Ser Ser Gly Gly Tyr
        245                 250                 255
Trp Val Gly Met Ser Thr Ile Arg Pro Asn Pro Gly Phe Ser Met Leu
260                 265                 270                 275
Asp Phe Leu Ser Glu Arg Pro Trp Ile Lys Arg Met Ile Phe Lys Ala
                280                 285                 290
Lys Lys Lys
```

<210> 236
<211> 106
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -32..-1

<400> 236
```
Met Phe Ala Pro Ala Val Met Arg Ala Phe Arg Lys Asn Lys Thr Leu
        -30                 -25                 -20
Gly Tyr Gly Val Pro Met Leu Leu Leu Ile Val Gly Gly Ser Phe Gly
        -15                 -10                 -5
Leu Arg Glu Phe Ser Gln Ile Arg Tyr Asp Ala Val Lys Ser Lys Met
1                   5                   10                  15
Asp Pro Glu Leu Glu Lys Lys Leu Lys Glu Asn Lys Ile Ser Leu Glu
                20                  25                  30
Ser Glu Tyr Glu Lys Ile Lys Asp Ser Lys Phe Asp Asp Trp Lys Asn
        35                  40                  45
Ile Arg Gly Pro Arg Pro Trp Glu Asp Pro Asp Leu Leu Gln Gly Arg
        50                  55                  60
Asn Pro Glu Ser Leu Lys Thr Lys Thr Thr
65                  70
```

<210> 237
<211> 42
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -19..-1

<400> 237
```
Met Asp Leu Arg Gln Phe Leu Met Cys Leu Ser Leu Cys Thr Ala Phe
                -15                 -10                 -5
Ala Leu Ser Lys Pro Thr Glu Lys Lys Asp Arg Val His His Glu Pro
```

```
                     1               5                          10
          Gln Leu Ser Asp Lys Val His Asn Asp Ile
            15                  20
```

<210> 238
<211> 117
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -20..-1

<400> 238
```
Met Asp Asn Arg Phe Ala Thr Ala Phe Val Ile Ala Cys Val Leu Ser
-20               -15                -10                      -5
Leu Ile Ser Thr Ile Tyr Met Ala Ala Ser Ile Gly Thr Asp Phe Trp
                 1               5                       10
Tyr Glu Tyr Arg Ser Pro Val Gln Glu Asn Ser Ser Asp Leu Asn Lys
            15                  20                  25
Ser Ile Trp Asp Glu Phe Ile Ser Asp Glu Ala Asp Glu Lys Thr Tyr
    30                  35                  40
Asn Asp Ala Leu Phe Arg Tyr Asn Gly Thr Val Gly Leu Trp Gly Arg
45                  50                  55                  60
Cys Ile Thr Ile Pro Lys Asn Met His Trp Tyr Ser Pro Pro Glu Arg
                 65                  70                  75
Thr Gly Ile Ser Leu Ile Leu Thr Ser Val Phe Phe Thr Trp Leu Ile
                 80                  85                  90
Ile Asp Lys Thr Thr
         95
```

<210> 239
<211> 178
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -37..-1

<400> 239
```
Met Glu Arg Gln Ser Arg Val Met Ser Glu Lys Asp Glu Tyr Gln Phe
         -35                  -30                  -25
Gln His Xaa Xaa Ala Xaa Xaa Leu Leu Val Phe Asn Phe Leu Leu Ile
    -20                  -15                  -10
Leu Thr Ile Leu Thr Ile Trp Leu Phe Lys Asn His Arg Phe Arg Phe
-5                   1               5                       10
Leu His Glu Thr Gly Gly Ala Met Val Tyr Gly Leu Ile Met Gly Leu
            15                  20                  25
Ile Ser Arg Tyr Ala Thr Ala Pro Thr Asp Ile Glu Ser Gly Thr Val
            30                  35                  40
Cys Asp Cys Val Lys Leu Thr Phe Ser Pro Pro Thr Leu Leu Val Asn
    45                  50                  55
Val Thr Asp Gln Val Tyr Glu Tyr Lys Tyr Lys Arg Glu Ile Ser Gln
60                  65                  70                  75
His Asn Ile Asn Pro His Gln Gly Asn Ala Ile Leu Glu Lys Met Thr
                 80                  85                  90
Phe Asp Pro Glu Ile Phe Phe Asn Val Leu Leu Pro Pro Ile Ile Phe
```

```
                95                    100                   105
His Ala Gly Tyr Ser Leu Lys Lys Arg His Phe Phe Gln Asn Leu Gly
        110                   115                   120
Ser Ile Leu Thr Tyr Ala Phe Leu Gly Thr Ala Ile Ser Cys Ile Val
        125                   130                   135
Ile Gly
140
```


```
<210> 240
<211> 126
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -27..-1

<400> 240
Met Gln Phe Val Asn Val Gly Tyr Phe Leu Ile Ala Ala Gly Val Val
        -25                   -20                   -15
Val Leu Ala Leu Gly Phe Leu Gly Cys Tyr Gly Ala Lys Thr Glu Ser
        -10                   -5                    1                     5
Met Cys Ala Leu Val Thr Phe Phe Phe Ile Leu Leu Leu Ile Phe Ile
                10                    15                    20
Ala Glu Val Ala Ala Ala Val Val Ala Leu Val Tyr Thr Thr Met Ala
                25                    30                    35
Glu His Phe Leu Thr Leu Leu Val Val Pro Ala Ile Lys Lys Asp Tyr
        40                    45                    50
Gly Ser Gln Glu Asp Phe Thr Gln Val Trp Asn Thr Thr Met Lys Gly
        55                    60                    65
Leu Lys Cys Arg Gly Phe Thr Asn Tyr Thr Asp Phe Glu Asp Ser Pro
70                    75                    80                    85
Tyr Phe Lys Met His Lys Pro Val Thr Met Lys Lys Lys Lys
                90                    95
```


```
<210> 241
<211> 174
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -115..-1

<400> 241
Met Arg Trp Ser Cys Glu His Leu Val Met Val Trp Ile Asn Ala Phe
-115                   -110                  -105                  -100
Val Met Leu Thr Thr Gln Leu Leu Pro Ser Lys Tyr Cys Asp Leu Leu
                -95                   -90                   -85
His Lys Ser Ala Ala His Leu Gly Lys Trp Gln Lys Leu Glu His Gly
                -80                   -75                   -70
Ser Tyr Ser Asn Ala Pro Gln His Ile Trp Ser Glu Asn Thr Ile Trp
        -65                   -60                   -55
Pro Gln Gly Val Leu Val Arg His Ser Arg Cys Leu Tyr Arg Ala Met
        -50                   -45                   -40
Gly Pro Tyr Asn Val Ala Val Pro Ser Asp Val Ser His Ala Arg Phe
-35                   -30                   -25                   -20
Tyr Phe Leu Phe His Arg Pro Leu Arg Leu Leu Asn Leu Leu Ile Leu
```

```
                      -15                    -10                     -5
          Ile Glu Gly Gly Val Val Phe Tyr Gln Leu Tyr Ser Leu Leu Arg Ser
                      1                     5                     10
          Glu Lys Trp Asn His Thr Leu Ser Met Ala Leu Ile Leu Phe Cys Asn
              15                    20                    25
          Tyr Tyr Val Leu Phe Lys Leu Leu Arg Asp Arg Ile Val Leu Gly Arg
          30                    35                    40                    45
          Ala Tyr Ser Tyr Pro Leu Asn Ser Tyr Glu Leu Lys Ala Asn
                      50                    55
```

<210> 242
<211> 896
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 18..173

<221> sig_peptide
<222> 18..77
<223> Von Heijne matrix
      score 6.5
      seq GLCVLQLTTAVTS/AF

<221> polyA_signal
<222> 864..869

<221> polyA_site
<222> 882..893

<400> 242
```
aaccttcaca gtgtgag atg cct agt gtg aac agt gct gga tta tgt gtc        50
                    Met Pro Ser Val Asn Ser Ala Gly Leu Cys Val
                    -20                    -15                   -10
ttg cag ttg aca acg gca gtr acc agt gcc ttt tta cta gca aaa gtg        98
Leu Gln Leu Thr Thr Ala Val Thr Ser Ala Phe Leu Leu Ala Lys Val
                -5                     1                     5
aat cct ttc gaa rct ttt ctc tca agg ggc ttt tgg cta tgt gct gcc      146
Asn Pro Phe Glu Xaa Phe Leu Ser Arg Gly Phe Trp Leu Cys Ala Ala
            10                    15                    20
cat cat ttc att cat cct tgc ctg gat tgagacgtgt tcctgattca            193
His His Phe Ile His Pro Cys Leu Asp
        25                    30
aagtgttacc tcaagaagca gaagaagaaa acagactcct gatagttcag gatgcttcag     253
agagggcagc acttatacct ggtggtcttt ctgatggtca gttttattcc cctcctgaat     313
ccgaagcagg atctgaagaa gctgaagaaa acaggacag tgagaaacca cttttagaac      373
tatgagtact actttgtta aatgtgaaaa accctcacag aaagtcatcg aggcaaaaag      433
aggcaggcag tggagtctcc ctgtcgacag taaagttgaa atggtacgt ccactgctgg      493
ctttattgaa cagctaataa agatttattt attgtaatac ctcacagacg ttgtaccata     553
tccatgcaca tttagttgcc tgcctgtggc tggtaaggta atgtcatgat tcatcctctc     613
ttcagtgaga ctgagcctga tgtgttaaca aataggtgaa gaaagtcttg tgctgtattc     673
ctaatcaaaa gacttaatat attgaagtaa cactttttta gtaagcaaga tacctttta     733
tttcaattca cagaatggaa ttttttttgtt tcatgtctca gatttatttt gtatttcttt    793
tttaacactc tacatttccc ttgtttttta actcatgcac atgtgctctt tgtacagttt     853
taaaaagtgt aataaaatct gacatgtcaa araaaaaaaa mcy                       896
```

<210> 243

<211> 851
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 17..595

<221> sig_peptide
<222> 17..85
<223> Von Heijne matrix
      score 3.70000004768372
      seq FLPPLXRAFACRG/CQ

<221> polyA_signal
<222> 820..825

<221> polyA_site
<222> 840..851

<400> 243
```
aaggggggcgt ggggcc atg gtg gtc ttg cgg gcg ggg aag aag acc ttt ctc       52
               Met Val Val Leu Arg Ala Gly Lys Lys Thr Phe Leu
                   -20                     -15
ccc cct ctm wgc cgc gcc ttc gcc tgc cgc ggc tgt caa ctc gct ccg         100
Pro Pro Leu Xaa Arg Ala Phe Ala Cys Arg Gly Cys Gln Leu Ala Pro
        -10             -5              1               5
gag cgc ggc gcc gag cgc agg gat aca gcg ccc agc ggg gtc tca aga         148
Glu Arg Gly Ala Glu Arg Arg Asp Thr Ala Pro Ser Gly Val Ser Arg
            10              15              20
ttc tgc cct cca aga aag tct tgc cat gat tgg ata gga ccc cca gat         196
Phe Cys Pro Pro Arg Lys Ser Cys His Asp Trp Ile Gly Pro Pro Asp
            25              30              35
aaa tat tca aac ctt cga cct gtt cac ttt tac ata cct gaa aat gaa         244
Lys Tyr Ser Asn Leu Arg Pro Val His Phe Tyr Ile Pro Glu Asn Glu
        40              45              50
tct cca ttg gaa caa aag ctt aga aaa tta aga caa gaa aca caa gaa         292
Ser Pro Leu Glu Gln Lys Leu Arg Lys Leu Arg Gln Glu Thr Gln Glu
        55              60              65
tgg aat caa cag ttc tgg gca aac cag aat ttg act ttt agt aag gaa         340
Trp Asn Gln Gln Phe Trp Ala Asn Gln Asn Leu Thr Phe Ser Lys Glu
70              75              80              85
aaa gaa gaa ttt att cac tca aga cta aaa act aaa ggc ctg ggc ctg         388
Lys Glu Glu Phe Ile His Ser Arg Leu Lys Thr Lys Gly Leu Gly Leu
                90              95              100
aga act gaa tca ggt cag aaa gca aca ttg aat gca gaa gaa atg gcg         436
Arg Thr Glu Ser Gly Gln Lys Ala Thr Leu Asn Ala Glu Glu Met Ala
            105             110             115
gac ttc tac aag gaa ttt tta agt aaa aat ttt cag aag cac atg tat         484
Asp Phe Tyr Lys Glu Phe Leu Ser Lys Asn Phe Gln Lys His Met Tyr
        120             125             130
tat aac aga gat tgg tac aag cgc aat ttt gcc atc acc ttc ttc atg         532
Tyr Asn Arg Asp Trp Tyr Lys Arg Asn Phe Ala Ile Thr Phe Phe Met
        135             140             145
gga aaa gtg gcc ctg gaa agg att tgg aac aag ctt aaa cag aaa caa         580
Gly Lys Val Ala Leu Glu Arg Ile Trp Asn Lys Leu Lys Gln Lys Gln
150             155             160             165
aag aag agg agc aac taggagtcca ctctgaccca gccagagtcc aggtttccac        635
Lys Lys Arg Ser Asn
                170
aggaagcara tggagctcct ttcacagggg ctctgagaaa aactggagct gatctcaaga      695
agccccacat cttcctaagg ggccccatgg cctgtttggg ggcagggtag gtcctggggc      755
```

```
actgtgggcc gcctgcctgc tgatgtgggc tctaggccag cttgttgtca cgtacgtggt    815
gtgaaataaa gcccaagcac tgggaaaaaa aaaaaa                              851
```

<210> 244
<211> 495
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 89..334

<221> sig_peptide
<222> 89..130
<223> Von Heijne matrix
      score 3.59999990463257
      seq AFTLXSLLQAALL/CV

<221> polyA_signal
<222> 462..467

<221> polyA_site
<222> 484..495

<400> 244
```
agtaggaasg cgccgsccgt ggaggcgcca cgtcccttgc sgcggcggga gagamatcgc    60
ttggacttcg gggcggcctc ggacggcc atg gcc ttt acc ctg tas tca ctg      112
                              Met Ala Phe Thr Leu Xaa Ser Leu
                                                  -10
ctg cag gca gcc ctg ctc tgc gtc aac gcc atc gca gtg ctg cac gag    160
Leu Gln Ala Ala Leu Leu Cys Val Asn Ala Ile Ala Val Leu His Glu
    -5                    1                5                10
gag cga ttc ctc aag aac att ggc tgg gga aca gac cag gga att ggt    208
Glu Arg Phe Leu Lys Asn Ile Gly Trp Gly Thr Asp Gln Gly Ile Gly
             15                20                25
gga ttt gga gaa gag ccg gga att aaa tca sag sta atg avs ctt att    256
Gly Phe Gly Glu Glu Pro Gly Ile Lys Ser Xaa Xaa Met Xaa Leu Ile
             30                35                40
cga tct gta aga acc gtg atg aga gtg cca ttg ata ata gta aac tca    304
Arg Ser Val Arg Thr Val Met Arg Val Pro Leu Ile Ile Val Asn Ser
             45                50                55
att gca att gtg tta ctt tta tta ttt gga tgaatwtcat ggagaaaat       354
Ile Ala Ile Val Leu Leu Leu Leu Phe Gly
             60                65
ggakactcag aaraggacat gccaktaraa kttattactt tggtcattat tggaatattt    414
atatcttagc tggctgacct tgcacttgtc aaaaatgtaa agctgaaaat aaaaccaggg    474
tttctattta aaaaaaaaaa a                                            495
```

<210> 245
<211> 884
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 21..614

<221> sig_peptide

```
<222>  21..83
<223>  Von Heijne matrix
       score 10
       seq LWALAMVTRPASA/AP

<221>  polyA_signal
<222>  849..854

<221>  polyA_site
<222>  873..884


<400>  245
aataccttag accctcagtc atg cca gtg cct gct ctg tgc ctg ctc tgg gcc        53
                      Met Pro Val Pro Ala Leu Cys Leu Leu Trp Ala
                      -20                     -15
ctg gca atg gtg acc cgg cct gcc tca gcg gcc ccc atg ggc ggc cca         101
Leu Ala Met Val Thr Arg Pro Ala Ser Ala Ala Pro Met Gly Gly Pro
-10                 -5                  1                   5
gaa ctg gca cag cat gag gag ctg acc ctg ctc ttc cat ggg acc ctg         149
Glu Leu Ala Gln His Glu Glu Leu Thr Leu Leu Phe His Gly Thr Leu
                10              15                  20
cag ctg ggc cag gcc ctc aac ggt gtg tac agg acc acg gag gga cgg         197
Gln Leu Gly Gln Ala Leu Asn Gly Val Tyr Arg Thr Thr Glu Gly Arg
            25              30              35
ctg aca aag gcc agg aac agc ctg ggt ctc tat ggc cgc aca ata gaa         245
Leu Thr Lys Ala Arg Asn Ser Leu Gly Leu Tyr Gly Arg Thr Ile Glu
        40              45              50
ctc ctg ggg cag gag gtc agc cgg ggc cgg gat gca gcc cag gaa ctt         293
Leu Leu Gly Gln Glu Val Ser Arg Gly Arg Asp Ala Ala Gln Glu Leu
55              60                  65                  70
cgg gca agc ctg ttg gaa act car atg gag gag gat att ctg cas ctg         341
Arg Ala Ser Leu Leu Glu Thr Gln Met Glu Glu Asp Ile Leu Xaa Leu
            75                  80                  85
cag gca rag gcc aca gct gag gtg ctg ggg gag gtg gcc cag gca car         389
Gln Ala Xaa Ala Thr Ala Glu Val Leu Gly Glu Val Ala Gln Ala Gln
            90              95                  100
aag gtg cta cgg gac agc gtg cag cgg cta daa ktc cag ctg arg asc         437
Lys Val Leu Arg Asp Ser Val Gln Arg Leu Xaa Xaa Gln Leu Xaa Xaa
                105             110                 115
gcc tgg ctg ggc cct gcc tac cga aaa ttt gar gtc tta aag gcy ccc         485
Ala Trp Leu Gly Pro Ala Tyr Arg Lys Phe Glu Val Leu Lys Ala Pro
    120                 125                 130
cck gam aar car aac cac atc cta tgg gcc ctc aca ggc cac gtg cak         533
Pro Xaa Lys Gln Asn His Ile Leu Trp Ala Leu Thr Gly His Val Xaa
135                 140                 145                 150
cgg car agg cgg gar atg gtg gca cag cag cwt ckg ctg cna car atc         581
Arg Gln Xaa Arg Glu Met Val Ala Gln Gln Xaa Xaa Leu Xaa Gln Ile
            155                 160                 165
cag gar aaa ctc cac aca gcg gcg ctc cca gcc tgaatctgcc tggatggaac       634
Gln Glu Lys Leu His Thr Ala Ala Leu Pro Ala
            170                 175
tgaggaccaa tcatgctgca aggaacactt ccacgccccg tgaggcccct gtgcagggag       694
gagctgcctg ttcactggga tcagccaggg cgccgggccc cacttctgag cacagagcar       754
agacagacgc aggcggggac aaaggcagag gatgtagccc cattggggag gggtggagga       814
aggacatgta cccttcatr mctacacacc cctcattaaa gcavagtcgt ggcatctcaa        874
aaaaaaaaaa                                                              884



<210>  246
<211>  897
<212>  DNA
```

```
<213> Homo sapiens

<220>
<221> CDS
<222> 94..573

<221> sig_peptide
<222> 94..258
<223> Von Heijne matrix
      score 4.69999980926514
      seq IGILCSLLGTVLL/WV

<221> polyA_signal
<222> 862..867

<221> polyA_site
<222> 886..897

<400> 246
aagggcggct gcctagcacc cggaagagcc gtcaacttag cgagcgcaac aggctgccgc      60
tgaggagctg gagctggtgg ggactgggcc gca atg gac aag ctg aag aag gtg     114
                                    Met Asp Lys Leu Lys Lys Val
                                    -55                 -50
ctg agc ggg cag gac acg gag gac cgg agc ggc ctg tcc gag gtt gtt     162
Leu Ser Gly Gln Asp Thr Glu Asp Arg Ser Gly Leu Ser Glu Val Val
        -45             -40             -35
gag gca tct tca tta agc tgg agt acc agg ata aaa ggc ttc att gcg     210
Glu Ala Ser Ser Leu Ser Trp Ser Thr Arg Ile Lys Gly Phe Ile Ala
        -30             -25             -20
tgt ttt gct ata gga att ctc tgc tca ctg ctg ggt act gtt ctg ctg     258
Cys Phe Ala Ile Gly Ile Leu Cys Ser Leu Leu Gly Thr Val Leu Leu
    -15             -10             -5
tgg gtg ccc agg aag gga cta cac ctc ttc gca gtg ttt tat acc ttt     306
Trp Val Pro Arg Lys Gly Leu His Leu Phe Ala Val Phe Tyr Thr Phe
1               5               10              15
ggt aat atc gca tca att ggg agt acc atc ttc ctc atg gga cca gtg     354
Gly Asn Ile Ala Ser Ile Gly Ser Thr Ile Phe Leu Met Gly Pro Val
            20              25              30
aaa cag ctg aag cga atg ttt gag cct act cgt ttg att gca act atc     402
Lys Gln Leu Lys Arg Met Phe Glu Pro Thr Arg Leu Ile Ala Thr Ile
        35              40              45
atg gtg ctg ttg tgt ttt gca ctt acc ctg tgt tct gcc ttt tgg tgg     450
Met Val Leu Leu Cys Phe Ala Leu Thr Leu Cys Ser Ala Phe Trp Trp
    50              55              60
cat aac aag gga ctt gca ctt atc ttc tgc att ttg cag tct ttg gca     498
His Asn Lys Gly Leu Ala Leu Ile Phe Cys Ile Leu Gln Ser Leu Ala
65              70              75              80
ttg acg tgg tac agc ctt tcc ttc ata cca ttt gca agg gat gct gtg     546
Leu Thr Trp Tyr Ser Leu Ser Phe Ile Pro Phe Ala Arg Asp Ala Val
            85              90              95
aaa aad tgt ttt gcc gtg tgt ctt gca taattcatgg ccagttttat          593
Lys Xaa Cys Phe Ala Val Cys Leu Ala
        100             105
gaagctttgg aaggcactat ggacagaagc tggtggacag ttttgtwact atcttcgaaa     653
cctctgtctt acagacatgt gcctttatc ttgcagcaat gtgttgcttg tgattcgaac     713
atttgagggt tactttggaa gcaacaata cattctcgaa cctgaatgtc agtagcacag     773
gatgagaagt gggttctgta tcttgtggag tggaatcttc ctcatgtacc tgtttcctct     833
ctggatgttg tcccactgaa ttcccatgaa tacaaaccta ttcagcaaca gcaaaaaaaa     893
aaaa                                                                  897
```

```
<210> 247
<211> 518
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 74..397

<221> sig_peptide
<222> 74..127
<223> Von Heijne matrix
      score 7.69999980926514
      seq LLLLPVLGLLVSS/KT

<221> polyA_signal
<222> 472..477

<221> polyA_site
<222> 507..518

<400> 247
aaagaaagag ctgcsgtgca ggaattcgtg tgccggattt ggttagctga gcccaccgag        60
aggcgcctgc agg atg aaa gct ctc tgt ctc ctc ctc ctc cct gtc ctg         109
                Met Lys Ala Leu Cys Leu Leu Leu Leu Pro Val Leu
                    -15                   -10
ggg ctg ttg gtg tct agc aag acc ctg tgc tcc atg gaa gaa gcc atc        157
Gly Leu Leu Val Ser Ser Lys Thr Leu Cys Ser Met Glu Glu Ala Ile
    -5                1                5                       10
aat gag agg atc cag gag gtc gcc ggc tcc cta ata ttt agg gca ata        205
Asn Glu Arg Ile Gln Glu Val Ala Gly Ser Leu Ile Phe Arg Ala Ile
                15                20                25
agc agc att ggc cga ggg agc gag agc gtc acc tcc agg ggg gac ctg        253
Ser Ser Ile Gly Arg Gly Ser Glu Ser Val Thr Ser Arg Gly Asp Leu
                30                35                40
gct act tgc ccc cga ggc ttc gcc gtc acc ggc tgc act tgt ggc tcc        301
Ala Thr Cys Pro Arg Gly Phe Ala Val Thr Gly Cys Thr Cys Gly Ser
                45                50                55
gcc tgt ggc tcg tgg gat gtg cgc gcc gag acc aca tgt cac tgc cag        349
Ala Cys Gly Ser Trp Asp Val Arg Ala Glu Thr Thr Cys His Cys Gln
    60                65                70
tgc gcg ggc atg gac tgg acc gga gcg cgc tgc tgt cgt gtg cag ccc        397
Cys Ala Gly Met Asp Trp Thr Gly Ala Arg Cys Cys Arg Val Gln Pro
75                80                85                90
tgaggtcgcg cgcagcgcgt gcacagcgcg ggcggaggcg gctccaggtc cggaggggtt        457
gcgggggagc tggaaataaa cctggagatg atgatgatga tgatgatgga aaaaaaaaaa        517
a                                                                        518


<210> 248
<211> 350
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 51..242

<221> sig_peptide
<222> 51..116
<223> Von Heijne matrix
```

```
                    score 6.5
                    seq SCLCPALFPGTSS/FI

<221> polyA_signal
<222> 319..324

<221> polyA_site
<222> 339..350

<400> 248
acgtcattcc aaaaccacac ccttgcaaag ctttgtactc cgcaccccag atg atc          56
                                                                Met Ile
tcc agg cag ctc aga tct ctt tcc tgc ctt tgc cct gca ctg ttc ccc        104
Ser Arg Gln Leu Arg Ser Leu Ser Cys Leu Cys Pro Ala Leu Phe Pro
-20               -15                 -10                   -5
ggt act tcc tcc ttt att gta gca ctc agc tcc cca gcc gat ctg tac        152
Gly Thr Ser Ser Phe Ile Val Ala Leu Ser Ser Pro Ala Asp Leu Tyr
                   1             5               10
atc cct cav agg cas cga tct gat gaa ttg gtt ttt gaa tcc car aaa        200
Ile Pro Xaa Arg Xaa Arg Ser Asp Glu Leu Val Phe Glu Ser Gln Lys
            15                  20                  25
ggg tct gcc atg gag ttg gca gtc atc acg gta rat ggc gta              242
Gly Ser Ala Met Glu Leu Ala Val Ile Thr Val Xaa Gly Val
        30              35              40
tgattttgct gaattttaaa taaaatgaaa accataaatt acatratgct tttattgach     302
cttgacmact ggcctaaata aaaaractct gactccaaaa aaaaaaaa                   350




<210> 249
<211> 996
<212> DNA
<213> Homo sapiens

<220> ·
<221> CDS
<222> 111..191

<221> sig_peptide
<222> 111..155
<223> Von Heijne matrix
      score 5.80000019073486
      seq FLXLMTLTTHVHS/SA

<221> polyA_signal
<222> 965..970

<221> polyA_site
<222> 986..996

<400> 249
atccgataca gaacatgcag taatgtggac tgcccaccag aagcaggtga tttccgagct     60
cagcaatgct cagctcataa tgatgtcaag caccatggcc agttttatga atg ggy        116
                                                            Met Gly
                                                            -15
ttc ctg wgt cta atg acc ctg aca acc cat gtt cac tca agt gcc aag       164
Phe Leu Xaa Leu Met Thr Leu Thr Thr His Val His Ser Ser Ala Lys
            -10                 -5                  1
cca aat gaa caa ccc tgg ttg ttg aac tagcacctaa ggtcttarat            211
Pro Asn Glu Gln Pro Trp Leu Leu Asn
        5               10
ggtacgcgtt gctatacaga atctttggat atgtgcatca gtggtttatg ccaaattgtt    271
```

```
ggctgcgatc accagctggg aagcaccgtc aaggaarata actgtggggt ctgcaacrga    331
natgggtcca cctgccggct ggtccgaggg cartataaat cccakctctc cgcaaccaaa    391
tcrgatgata ctgtggttgc aattccctat ggaagtakac atattcgcct tgtcttaaaa    451
ggtcctgatc acttatatct ggaarccawa accctccagg ggactaawgg tgaaaacagt    511
ctcasctcca caggaacttt ccttgtggac aattctagtg tggacttcca gaawtttcca    571
gacwdagaga tactgagaat ggctggacca ctcacagcag atttcattgt caawattcgt    631
aactcgggct ccgctgacag tacagtccag kkcatcttct atcaacccat catccaccga    691
tggagggara cggatttctt tccttgctca gcaacctgtg gaggaggtta tcagctgaca    751
tcggctgagt gctacgatct gaggagcaac cgtgtggttg ctgaccaata ctgtcactat    811
tacccagaga acatcaaacc caaacccaag cttcaggagt gcaacttgga tccttgtcca    871
gccaggtcag tcaaatttgc tagttcattt gtcataaaca taactcaagt tccaaatagg    931
ttatttaaat taaaatgaaa cgttttaatt aaaaataaaa tgaaattaaa catcaaaaaa    991
aaaaa                                                                996
```

<210> 250
<211> 86C
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 45..602

<221> sig_peptide
<222> 45..107
<223> Von Heijne matrix
      score 8.5
      seq LLTIVGLILPTRG/QT

<221> polyA_signal
<222> 828..833

<221> polyA_site
<222> 850..860

<400> 250
```
acctctctcc acgaggctgc cggcttagga cccccagctc cgac atg tcg ccc tct    56
                                                Met Ser Pro Ser
                                                     -20
ggt cgc ctg tgt ctt ctc acc atc gtt ggc ctg att ctc ccc acc aga    104
Gly Arg Leu Cys Leu Leu Thr Ile Val Gly Leu Ile Leu Pro Thr Arg
         -15                -10                 -5
gga cag acg ttg aaa gat acc acg tcc agt tct tca gca gac tca act    152
Gly Gln Thr Leu Lys Asp Thr Thr Ser Ser Ser Ser Ala Asp Ser Thr
 1               5                   10                  15
atc atg gac att cag gtc ccg aca cga gcc cca gat gca gtc tac aca    200
Ile Met Asp Ile Gln Val Pro Thr Arg Ala Pro Asp Ala Val Tyr Thr
                20                  25                  30
gaa ctc cag ccc acc tct cca acc cca acc tgg cct gct gat gaa aca    248
Glu Leu Gln Pro Thr Ser Pro Thr Pro Thr Trp Pro Ala Asp Glu Thr
            35                  40                  45
cca caa ccc cag acc cag acc cag caa ctg gaa gga acg gat ggg cct    296
Pro Gln Pro Gln Thr Gln Thr Gln Gln Leu Glu Gly Thr Asp Gly Pro
        50                  55                  60
cta gtg aca gat cca gag aca cac wak agc mcc aaa gca gct cat ccc    344
Leu Val Thr Asp Pro Glu Thr His Xaa Ser Xaa Lys Ala Ala His Pro
    65                  70                  75
act gat gac acc acg acg ctc tct gag aga cca tcc cca agc aca kac    392
Thr Asp Asp Thr Thr Thr Leu Ser Glu Arg Pro Ser Pro Ser Thr Xaa
80                  85                  90                  95
```

```
gtc cat dac aga ccb cba kda ccc tca akc cat ctg gtt ttc atg agg      440
Val His Xaa Arg Pro Xaa Xaa Pro Ser Xaa His Leu Val Phe Met Arg
            100              105                  110
atg acc cct tct tct atg atg aac aca ccc tcc gga aac sgg ggc tgt      488
Met Thr Pro Ser Ser Met Met Asn Thr Pro Ser Gly Asn Xaa Gly Cys
            115              120                  125
tgg tcg cag ctg tgc tgt tca tca cag gca tca tca tcc tca cca gtg      536
Trp Ser Gln Leu Cys Cys Ser Ser Gln Ala Ser Ser Ser Ser Pro Val
            130              135                  140
gca agt gca ggc agc tgt ccc ggt tat gcc gga atc att gca ggt gag      584
Ala Ser Ala Gly Ser Cys Pro Gly Tyr Ala Gly Ile Ile Ala Gly Glu
            145              150                  155
tcc atc aga aac agg agc tgacaacctg ctgggcaccc gaagaccaag            632
Ser Ile Arg Asn Arg Ser
160              165
ccccctgcca gctcaccgtg cccagcctcc tgcatcccct cgaagagcct ggccagagag   692
ggaagacaca gatgatgaag ctggagccag ggctgccggt ccgagtctcc tacctccccc   752
aucccctgccc gcccctgaag gctacctggc gccttggggg ctgtccctca agttatctcc   812
tctgctaaga caaaaagtaa agcactgtgg tctttgcaaa aaaaaaaa                 860
```

```
<210> 251
<211> 593
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 24..560

<221> sig_peptide
<222> 24..101
<223> Von Heijne matrix
      score 10.3999996185303
      seq LLLLLCGPSQDQC/RP

<221> polyA_signal
<222> 563..568

<221> polyA_site
<222> 583..593

<400> 251
aanccagctg csgccggcca gcc atg gag act gga gcg ctg cgg cgc ccg caa   53
                        Met Glu Thr Gly Ala Leu Arg Arg Pro Gln
                        -25                      -20
ctt ctc ccg ttg ctg ctg ctc ctc tgc ggc cct tcc cag gat caa tgc   101
Leu Leu Pro Leu Leu Leu Leu Leu Cys Gly Pro Ser Gln Asp Gln Cys
    -15              -10                    -5
cga cct gta ctc cag aat ctg ttg cag agc cca ggc ttg aca tgg agc   149
Arg Pro Val Leu Gln Asn Leu Leu Gln Ser Pro Gly Leu Thr Trp Ser
1            5                    10                    15
ttg gaa gtg ccc act ggg aga gaa gga aag gaa ggt ggg gat cgg gga   197
Leu Glu Val Pro Thr Gly Arg Glu Gly Lys Glu Gly Gly Asp Arg Gly
            20               25                   30
cca ggg cta akt ggg gcc act cca gcc agg agc cct cag ggc aag gag   245
Pro Gly Leu Xaa Gly Ala Thr Pro Ala Arg Ser Pro Gln Gly Lys Glu
            35               40                   45
atg ggg aga caa agg acc aga aag gtg aag ggc cct gct tgg akt cac   293
Met Gly Arg Gln Arg Thr Arg Lys Val Lys Gly Pro Ala Trp Xaa His
    50               55                   60
```

```
aca gca aat cag gaa cta aac agg atg agg tct ctg tct tct ggc tcc        341
Thr Ala Asn Gln Glu Leu Asn Arg Met Arg Ser Leu Ser Ser Gly Ser
65                      70                  75                  80
gtg cca gtg ggg cat ctg gag ggt ggc acg gtc aag ctt cag aag gac        389
Val Pro Val Gly His Leu Glu Gly Gly Thr Val Lys Leu Gln Lys Asp
                    85                  90                  95
acg ggc ctc cat tcc tgc ara gat ggt atg gct tct ctt gaa ggg acg        437
Thr Gly Leu His Ser Cys Xaa Asp Gly Met Ala Ser Leu Glu Gly Thr
                100                 105                 110
cca gct tca gtc ctg gct gat gct tgc cca gga ttc cat gat gtg aan        485
Pro Ala Ser Val Leu Ala Asp Ala Cys Pro Gly Phe His Asp Val Xaa
            115                 120                 125
gtt car arg gcc cta ttt ggg tta agt ggg ana rta ctg tgg ctg aaa        533
Val Gln Xaa Ala Leu Phe Gly Leu Ser Gly Xaa Xaa Leu Trp Leu Lys
        130                 135                 140
acc cac ttc tgc ctt tct att ana ctt taaataaact ctgaaracct            580
Thr His Phe Cys Leu Ser Ile Xaa Leu
145                 150
gtaaaaaaaa aaa                                                          593
```

```
<210> 252
<211> 1114
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 109..558

<221> sig_peptide
<222> 109..273
<223> Von Heijne matrix
      score 3.70000004768372
      seq VAFMLTLPILVCK/VQ

<221> polyA_site
<222> 1104..1114

<400> 252
attagctstc caaggtctcc cccagcactg aggagctcgc ctgctgccct cttgcgcgcg        60
ggaagcagca ccaagttcac ggccaacgcc ttggcactag ggtccaga atg gct aca      117
                                                      Met Ala Thr
                                                          -55
aca gtc cct gat ggt tgc cgc aat ggc ctg aaa tcc aag tac tac aga        165
Thr Val Pro Asp Gly Cys Arg Asn Gly Leu Lys Ser Lys Tyr Tyr Arg
        -50                 -45                 -40
ctt tgt gat aag gct gaa gct tgg ggc atc gtc cta gaa acg gtg gcc        213
Leu Cys Asp Lys Ala Glu Ala Trp Gly Ile Val Leu Glu Thr Val Ala
    -35                 -30                 -25
aca gcc ggg gtt gtg acc tcg gtg gcc ttc atg ctg act ctc ccg atc        261
Thr Ala Gly Val Val Thr Ser Val Ala Phe Met Leu Thr Leu Pro Ile
-20                 -15                 -10                 -5
ctc gtc tgc aag gtg cag gac tcc aac agg cga aaa atg ctg cct act        309
Leu Val Cys Lys Val Gln Asp Ser Asn Arg Arg Lys Met Leu Pro Thr
                1                   5                   10
cag ttt ctc ttc ctc ctg ggt gtg ttg ggc atc ttt ggc ctc acc ttc        357
Gln Phe Leu Phe Leu Leu Gly Val Leu Gly Ile Phe Gly Leu Thr Phe
        15                  20                  25
gcc ttc atc atc gga ctg gac ggg agc aca ggg ccc aca cgc ttc ttc        405
Ala Phe Ile Ile Gly Leu Asp Gly Ser Thr Gly Pro Thr Arg Phe Phe
```

```
      30                  35                  40
ctc ttt ggg atc ctc ttt tcc atc tgc ttc tcc tgc ctg ctg gct cat    453
Leu Phe Gly Ile Leu Phe Ser Ile Cys Phe Ser Cys Leu Leu Ala His
45                  50                  55                  60
gct gtc agt ctg acc aag ctc gtc cgg ggg agg aaa gcc cct ttc cct    501
Ala Val Ser Leu Thr Lys Leu Val Arg Gly Arg Lys Ala Pro Phe Pro
                65                  70                  75
gtt ggt gat tct ggg tct ggc cgt ggg ctt cag cct agt cca gga tgt    549
Val Gly Asp Ser Gly Ser Gly Arg Gly Leu Gln Pro Ser Pro Gly Cys
            80                  85                  90
tat cgc tat tgaatatatt gtcctgacca tgaataggac caacgtcaat           598
Tyr Arg Tyr
        95
gtcttttctg agctttccgc tcctgtcgc aatgaaaact ttgtcctcct gctcacctac   658
ktcctcttct tgatggcgct gaccttcctc wtgtcctcct tcaccttctg tggtkccttc   718
acgggctgga avagacatgg ggcccacatc tacctcasga tgctcskctc cattgccatc   778
tgggtggcct ggatcaccct gctcatgctt cctgactttg accgcrggtg ggatgacacc   838
atcmtcarct ccgccttggs trcsaatggc tgsgtgttcc tgttggctta tgttagtccc   898
gagttttggc tgctcacaaa gcaackaaac cccatggatt atcctgttga ggatgctttc   958
tgtaaacctc aactcgtgaa gaagagctat ggtgtggrga acagagccta skctcaagag   1018
gaaatcactc aaggttttga agagacaggg gacacgctct atgcccccta ttccacacat   1078
tttcagctgc agaascagcc tccccaaaaa aaaaaa                            1114
```

<210> 253
<211> 1182
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 128..835

<221> sig_peptide
<222> 128..220
<223> Von Heijne matrix
      score 4.69999980926514
      seq LAVDSWWLDPGHA/AV

<221> polyA_signal
<222> 1145..1150

<221> polyA_site
<222> 1170..1181

<400> 253
```
aagaactgcg tctcgcgacc caggcgcggg ttcccggagg acagccaaca agcgatgctg    60
ccgccgccgt ttcctgattg gttgtgggtg gctacctctt cgttctgatt ggccgctagt   120
gagcaag atg ctg agc aag ggt ctg aag·cgg aaa cgg gag gag gag gag    169
        Met Leu Ser Lys Gly Leu Lys Arg Lys Arg Glu Glu Glu Glu
            -30                 -25                 -20
gag aag gaa cct ctg gca gtc gac tcc·tgg tgg cta gat cct ggc cac    217
Glu Lys Glu Pro Leu Ala Val Asp Ser Trp Trp Leu Asp Pro Gly His
        -15                 -10                 -5
gca gcg gtg gca cag gca ccc ccg gcc gtg gcc tct agc tcc ctc ttt    265
Ala Ala Val Ala Gln Ala Pro Pro Ala Val Ala Ser Ser Ser Leu Phe
    1                   5                   10                  15
gac ctc tca gtg ctc aag ctc cac cac agc ctg cag vrr agt rag ccg    313
Asp Leu Ser Val Leu Lys Leu His His Ser Leu Gln Xaa Ser Xaa Pro
                    20                  25                  30
gac ctg cgg cac ctg gtg ctg gtc atr aac act ctg cgg cgc atc cag    361
```

```
Asp Leu Arg His Leu Val Leu Val Xaa Asn Thr Leu Arg Arg Ile Gln
            35              40                      45
gcg tcc atg gca ccc gcg gct gcc ctg cca cct gtg cct acc cca cct      409
Ala Ser Met Ala Pro Ala Ala Ala Leu Pro Pro Val Pro Thr Pro Pro
        50                      55                      60
gca gcc ccc ant gtg gct gac aac tta ctg gca agc tcg gac gct gcc      457
Ala Ala Pro Xaa Val Ala Asp Asn Leu Leu Ala Ser Ser Asp Ala Ala
        65                      70                      75
ctt tca gcc tcc atg gcc arm ctc ctg gar gac ctc agc cac att gag      505
Leu Ser Ala Ser Met Ala Xaa Leu Leu Glu Asp Leu Ser His Ile Glu
80                      85                      90                  95
ggc ctg agt cag gct ccc caa ccc ttg gca gac gag ggg cca cca ggc      553
Gly Leu Ser Gln Ala Pro Gln Pro Leu Ala Asp Glu Gly Pro Pro Gly
                    100                     105                     110
cgt agc atc ggg gga wca ccg ccc amc ctg ggt gcc ttg gac ctg ctg      601
Arg Ser Ile Gly Gly Xaa Pro Pro Xaa Leu Gly Ala Leu Asp Leu Leu
             115                    120                     125
ggc cca gcc act ggc tgt cta ctg gac aat ggg ctt gag ggc ctg ttt      649
Gly Pro Ala Thr Gly Cys Leu Leu Asp Asn Gly Leu Glu Gly Leu Phe
            130                     135                     140
gag gat att gac acc tct atg tat gac aat gaa ctt tgg gca cca gcc      697
Glu Asp Ile Asp Thr Ser Met Tyr Asp Asn Glu Leu Trp Ala Pro Ala
            145                     150                     155
tct gag ggc ctc aaa cca ggc cct gag gat ggg ccg ggc aag gag gaa      745
Ser Glu Gly Leu Lys Pro Gly Pro Glu Asp Gly Pro Gly Lys Glu Glu
160                     165                     170                 175
gct ccg gag ctg gac gag gcc gaa ttg gac tac ctc atg gat gtg ctg      793
Ala Pro Glu Leu Asp Glu Ala Glu Leu Asp Tyr Leu Met Asp Val Leu
                    180                     185                     190
gtg ggc aca cag gca ctg gag cga ccg ccg ggg cca ggg cgc              835
Val Gly Thr Gln Ala Leu Glu Arg Pro Pro Gly Pro Gly Arg
                195                     200                     205
tgagccctcg tgctggaatg gttgtctggt atctgaactg agcctgctgg ctggaccaac     895
tgtcctcgaa aagacacagc tggcttccct agtacagaga acagggcttg ggccactttg     955
gagagacaga atctagtcct gggcaacttc acatccgtcc tcctgtctca gggctggcag    1015
ggggagcctg gaattacccc ctagtgatgg aatgacaggg tctggtgggg actgaattcc    1075
ctggccctgg ggtcatagct tgggctgttc cttctctgat acgggaagag acccaatcag    1135
attttttcaaa ttaaagccag tcctgggaaa tctcaaaaaa aaaaaac               1182
```

```
<210> 254
<211> 1073
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 59..505

<221> sig_peptide
<222> 59..358
<223> Von Heijne matrix
      score 3.70000004768372
      seq LASSFLFTMGGLG/FI

<221> polyA_signal
<222> 1042..1047

<221> polyA_site
<222> 1062..1073
```

```
<400> 254
actgtttnng ggaggcgcgt ggggcttgag gccgagaacg gcccttgctg ccaccaac        58
atg gag act ttg tac cgt gtc ccg ttc tta gtg ctc gaa tgt ccc aac       106
Met Glu Thr Leu Tyr Arg Val Pro Phe Leu Val Leu Glu Cys Pro Asn
-100                -95                 -90                 -85
ctg aag ctg aag aag ccg ccc tgg ttg cac atg ccg tcg gcc atg act       154
Leu Lys Leu Lys Lys Pro Pro Trp Leu His Met Pro Ser Ala Met Thr
            -80                 -75                 -70
gtg tat gct ctg gtg gtg gtg tct tac ttc ctc atc acc gga gga ata       202
Val Tyr Ala Leu Val Val Val Ser Tyr Phe Leu Ile Thr Gly Gly Ile
        -65                 -60                 -55
att tat gat gtt att gtt gaa cct cca agt gtc ggt tct atg act gat       250
Ile Tyr Asp Val Ile Val Glu Pro Pro Ser Val Gly Ser Met Thr Asp
        -50                 -45                 -40
gaa cat ggg cat cag agg cca gta gct ttc ttg gcc tac aga gta aat       298
Glu His Gly His Gln Arg Pro Val Ala Phe Leu Ala Tyr Arg Val Asn
        -35                 -30                 -25
gga caa tat att atg gaa gga ctt gca tcc agc ttc cta ttt aca atg       346
Gly Gln Tyr Ile Met Glu Gly Leu Ala Ser Ser Phe Leu Phe Thr Met
-20                 -15                 -10                 -5
gga ggt tta ggt ttc ata atc ctg gac gga tcg aat gca cca aat atc       394
Gly Gly Leu Gly Phe Ile Ile Leu Asp Gly Ser Asn Ala Pro Asn Ile
            1                   5                   10
cca aaa ctc aat aga ttc ctt ctt ctg ttc att gga ttc gtc tgt gtc       442
Pro Lys Leu Asn Arg Phe Leu Leu Leu Phe Ile Gly Phe Val Cys Val
        15                  20                  25
cta twr agt ttt tkc ayg gct aga gta ttc atg aga atg aaa ctg ccg       490
Leu Xaa Ser Phe Xaa Xaa Ala Arg Val Phe Met Arg Met Lys Leu Pro
        30                  35                  40
ggc tat ctg atg ggt tagagtgcct ttgasaagaa atcagtggat actggatttg       545
Gly Tyr Leu Met Gly
45
ctcctgtcaa wgaastttta aaggctgtmc caatcctcta atatgaaatg tggaaaagaa       605
tgaagagcag cagtaaaaga aatatctagt gaaaaaacag gaagcgtatt gaagcttgga       665
ctagaatttc ttcttggtat taaagagaca agtttatcac agaatttttt ttcctgctgg       725
cctattgcta taccaatgat gttgagtggc attttctttt tagttttttca ttaaaatata       785
ttccatatct acaactataa tatcaaataa agtgattatt ttttacaacc ctcttaacat       845
tttttggaga tgacatttct gattttcaga aattaacata aaatccagaa gcaagattcc       905
gtaagctgag aactctggac agttgatcag ctttacctat ggtgctttgc ctttaactag       965
agtgtgtgat ggtagattat ttcagatatg tatgtaaaac tgtttcctga acaataagat      1025
gtatgaacgg agcagaaata aatacttttt ctaattaaaa aaaaaaaa                   1073
```

```
<210> 255
<211> 818
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 1..207

<221> sig_peptide
<222> 1..147
<223> Von Heijne matrix
      score 7.59999990463257
      seq HLPFLLLLSCVGX/XP

<221> polyA_signal
<222> 784..789
```

```
<221> polyA_site
<222> 807..818


<400> 255
atg cct ttc cat ttt ccg ttc ctt ggg ttt gtg tgt ctg cat ctc cat     48
Met Pro Phe His Phe Pro Phe Leu Gly Phe Val Cys Leu His Leu His
              -45                 -40                 -35
ctt acc cct tgc ctg act gta ccc cgt aga ccc ctg ttt ctc ctc ctg     96
Leu Thr Pro Cys Leu Thr Val Pro Arg Arg Pro Leu Phe Leu Leu Leu
          -30                 -25                 -20
cac ctg tgt ccc cat ctg ccc ttc ttg ttg ctc ctg tca tgt gtc ggg    144
His Leu Cys Pro His Leu Pro Phe Leu Leu Leu Leu Ser Cys Val Gly
          -15                 -10                  -5
gkc www ccc tcc tgt ctg cct tct tcc tcc act tgt gtc agc ttg cat    192
Xaa Xaa Pro Ser Cys Leu Pro Ser Ser Ser Thr Cys Val Ser Leu His
    1               5                  10                  15
ttt ttt att cct gac tgagtcacca caccctctc ccctgatcaa agggaatatk    247
Phe Phe Ile Pro Asp
            20
arttttttaat ttggatcgac tgaggtgcca ggagaaactg cagkcccagg tatccmvaca  307
gccaccagga tggtccctcg ccccacccc accgcctctk ccccacctтt tccaacgtgt    367
tgcatgctgg gaactggggg gtgtgggggga aggggctgcc ggcttctttc aggangctga  427
rgtttggarg caaaatcaac ctgggaracc accccggccg cggcgcctca gtggacaggt    487
gggargaaaa gaaaacttct taccttggar garggacatc ccgcttcctt atccttagct    547
tttttgttgc tcctcccac tgcccctttt aatttatttg gttgtttgcg gaaggagggg    607
ggaaggggt aagctgggcc gggaactgtc cgaggtgctg agctggggcg ggaccggaat    667
cctcccggta gggtaccagg gactgagttg ggcctggggc cgtgtccaag gtgccaatga    727
tgcgggccga cagarcgggc cgcactgtct gtctgtccgt ctgtcccgga aagaactata    787
aagcgctgga agcgcctgca aaaaaaaaa a                                     818




<210> 256
<211> 971
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 12..734

<221> sig_peptide
<222> 12..101
<223> Von Heijne matrix
      score 4.80000019073486
      seq ILFCVGAVGACTL/SV

<221> polyA_signal
<222> 914..919

<221> polyA_site
<222> 961..971


<400> 256
aatacacaga a atg ggg act gcg agc aga agc aac atc gct cgc cat ctg     50
            Met Gly Thr Ala Ser Arg Ser Asn Ile Ala Arg His Leu
            -30                 -25                 -20
caa acc aat ctc att cta ttt tgt gtc ggt gct gtg ggc gcc tgt act     98
Gln Thr Asn Leu Ile Leu Phe Cys Val Gly Ala Val Gly Ala Cys Thr
          -15                 -10                  -5
ctc tct gtc aca caa ccg tgg tac cta gaa gtg gac tac act cat gag    146
Leu Ser Val Thr Gln Pro Trp Tyr Leu Glu Val Asp Tyr Thr His Glu
```

```
      1             5                    10                 15
gcc gtc acc ata aag tgt acc ttc tcc gca acc gga tgc cct tct gag        194
Ala Val Thr Ile Lys Cys Thr Phe Ser Ala Thr Gly Cys Pro Ser Glu
                  20                   25                 30
caa cca aca tgc ctg tgg ttt cgc tac ggt gct cac cag cct gag aac        242
Gln Pro Thr Cys Leu Trp Phe Arg Tyr Gly Ala His Gln Pro Glu Asn
                  35                   40                 45
ctg tgc ttg gac ggg tgc aaa agt gag gca gas aag ttc aca gtg agg        290
Leu Cys Leu Asp Gly Cys Lys Ser Glu Ala Xaa Lys Phe Thr Val Arg
                  50                   55                 60
gag gcc ctc aaa gaa aac caa gtt tcc ctc act gta aac aga gtg act        338
Glu Ala Leu Lys Glu Asn Gln Val Ser Leu Thr Val Asn Arg Val Thr
          65                   70                   75
tca aat gac agt gca att tac atc tgt gga ata gca ttc ccc agt gtg        386
Ser Asn Asp Ser Ala Ile Tyr Ile Cys Gly Ile Ala Phe Pro Ser Val
80                   85                   90                 95
ccg gaa gcg aga gct aaa cag aca gga gga ggg acc aca ctg gtg gta        434
Prc Glu Ala Arg Ala Lys Gln Thr Gly Gly Gly Thr Thr Leu Val Val
                  100                  105                110
aga gaa att aag ctg ctc agc aag gaa ctg cgg agc ttc ctg aca gct        482
Arg Glu Ile Lys Leu Leu Ser Lys Glu Leu Arg Ser Phe Leu Thr Ala
              115                  120                  125
ctt gta tca ctg ctc tct gtc tat gtg acc ggt gtg tgc gtg gcc ttc        530
Leu Val Ser Leu Leu Ser Val Tyr Val Thr Gly Val Cys Val Ala Phe
              130                  135                  140
ata ctc ctc tcc aaa tca aaa tcc aac cct cta aga aac aaa gaa ata        578
Ile Leu Leu Ser Lys Ser Lys Ser Asn Pro Leu Arg Asn Lys Glu Ile
          145                  150                  155
aaa gaa gac tca caa aag aag aag agt gct cgg cgt att ttt cag gaa        626
Lys Glu Asp Ser Gln Lys Lys Lys Ser Ala Arg Arg Ile Phe Gln Glu
160                  165                  170                175
att gct caa gaa cta tac cat aag aga cat gtg gaa aca aat cag caa        674
Ile Ala Gln Glu Leu Tyr His Lys Arg His Val Glu Thr Asn Gln Gln
                  180                  185                  190
tct gag aaa gat aac aac act tat gaa aac aga aga gta ctt tcc aac        722
Ser Glu Lys Asp Asn Asn Thr Tyr Glu Asn Arg Arg Val Leu Ser Asn
              195                  200                  205
tat gaa agg cca tagaaacgtt ttaattttca atgaagtcac tgaaaatcca            774
Tyr Glu Arg Pro
              210
actccaggag ctatggcagt gttaatgaac atatatcatc aggtcttaaa aaaaaataaa      834
ggtaaactga aaagacaact ggctacaaag aaggatgcca raatgtaagg aaactataac      894
taataktcat taccaaaata ctaaaaccca acaaaatgca actgaaaaat accttccaaa      954
tttgccaaaa aaaaaaw                                                     971
```

<210> 257
<211> 640
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 378..518

<221> sig_peptide
<222> 378..467
<223> Von Heijne matrix
      score 5.5
      seq SLMTCTTLINASA/IS

```
<221> polyA_signal
<222> 607..612

<221> polyA_site
<222> 628..640

<400> 257
agcctgggta akgcccaaga tggctgtctt cgccttagta ctcgtgtgaa gttggcgggg    60
acggttcctg tcatcttctt gggcttattt ggtgtgctgt tgaagggggg agactagaga   120
aatggcaggg aacctcttat ccggggcagg taggcgcctg tgggactggg tgcctctggc   180
gtgcagaagc ttctctcttg gtgtgcctag attgatcggt ataaggctca ctctcccgcc   240
ccccaaagtg gttgatcgtt ggaacgagaa aagggccatg ttcggagtgt atgacaacat   300
cgggatcctg ggaaactttg aaaagcaccc caaagaactg atcaggggggc ccatatggct   360
tcgaggttgg aaaggga atg aat tgc aac gtt gta tcc gaa aga gga aaa       410
                     Met Asn Cys Asn Val Val Ser Glu Arg Gly Lys
                     -30              -25              -20
tgg ttg gaa gta gaa tgt tcg ctg atg acc tgc aca acc tta ata aac     458
Trp Leu Glu Val Glu Cys Ser Leu Met Thr Cys Thr Thr Leu Ile Asn
       -15              -10              -5
gca tcc gct atc tct aca aac act tta acc gac atg gga agt ttc gat     506
Ala Ser Ala Ile Ser Thr Asn Thr Leu Thr Asp Met Gly Ser Phe Asp
        1               5                10
aga aga gaa agc tgagaacttc ggaaaaggct catctgtcac cctggaraag         558
Arg Arg Glu Ser
        15
ggaaactgta cttttccctg tgaggaaacg gctttgtatt ttctctgtaa taaaatgggg   618
cttctttgga aaaaaaaaaa aa                                            640




<210> 258
<211> 745
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 110..304

<221> sig_peptide
<222> 110..193
<223> Von Heijne matrix
      score 4.59999990463257
      seq PLQWSLLVAVVAG/SV

<221> polyA_signal
<222> 708..713

<221> polyA_site
<222> 732..743

<400> 258
acttccgcct gcgcctgcgc agcvcagctc cshgagccct gccaaccatg gtgaacttgg    60
gtctgtcccg ggtggacgac gccgtggctg ccaagcaccc ggcaccggc atg gcc ttt   118
                                                      Met Ala Phe
ggc ttg cag atg ttc att cag agg aag ttt cca tac cct ttg cag tgg     166
Gly Leu Gln Met Phe Ile Gln Arg Lys Phe Pro Tyr Pro Leu Gln Trp
-25              -20              -15              -10
agc ctc cta gtg gcc gtg gtt gca ggc tct gtg gtc agc tac ggg gtg     214
Ser Leu Leu Val Ala Val Val Ala Gly Ser Val Val Ser Tyr Gly Val
        -5               1                5
acg aga gtg gag tcg gag aaa tgc aac aac ctc tgg ctc ttc ctg gag     262
```

```
Thr Arg Val Glu Ser Glu Lys Cys Asn Asn Leu Trp Leu Phe Leu Glu
        10              15              20
acc gga cag ctc ccc aaa gac agg agc aca gat cag ara agc          304
Thr Gly Gln Leu Pro Lys Asp Arg Ser Thr Asp Gln Xaa Ser .
      25              30              35
taggagagct ccagcagggg cacagargat tgggggcagg argartctgg aacacakcct  364
tcatgccccc tgaccccagg ccgaccctcc ccacaccta gggtacccca gtcgtatcct   424
ctgtccgcat gtgtggccag gcctgacaaa cmcctgcaga tggctgctgc cccaacctgg   484
gacctgccca ggaggttgga gcagaaaggg ctctccctgg ggtggtgttt ctcctctagg   544
gtattgggat gcatgttctg cactgccagc agagagggtg tgtctggggg ccaccaccta   604
tgggacacgg ggtcgaaggg gcctgtacac tctgtcattt ccttttctagc ccctgcatct   664
ccaacaagtc caaggtgaca gctggtgcta ggggcgtggg gttaataaat ggcttatcct   724
tctctccaaa araaaaaaam c                                           745
```

<210> 259
<211> 637
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 201..419

<221> sig_peptide
<222> 201..272
<223> Von Heijne matrix
      score 6.40000009536743
      seq LSYLPLWLGPIWP/CS

<221> polyA_signal
<222> 601..606

<221> polyA_site
<222> 627..637

```
<400> 259
acaaaatata attgcctcts ccctctccca ttttctctct tgggagcaat ggtcacagtc   60
cctggtacct gaaaaggtac ctaggtctag gcccttcttc cctttccctt cctctcccct  120
accccagaac tttggctccc ttcccttct ctctctggta gctccaggag gcctgtgatc   180
cagctccctg cctagcatcc atg acc tgt tgg atg tta cct cca atc agt ttc  233
                        Met Thr Cys Trp Met Leu Pro Pro Ile Ser Phe
                        -20                     -15
ctg tcc tac ctg cct ctt tgg ctt gga cct ata tgg cca tgc tct ggc    281
Leu Ser Tyr Leu Pro Leu Trp Leu Gly Pro Ile Trp Pro Cys Ser Gly
        -10             -5              1
tct acc ctt ggg aag cct gat ccc ggt gtg tgg ccc agc ttg ttc agg    329
Ser Thr Leu Gly Lys Pro Asp Pro Gly Val Trp Pro Ser Leu Phe Arg
      5               10              15
ccc tgg gat gct gca tct cca ggc aac tat gca ctt tcc cgg gga rar    377
Pro Trp Asp Ala Ala Ser Pro Gly Asn Tyr Ala Leu Ser Arg Gly Xaa
20              25              30              35
aac cak tat gav aak tgg ggg cag ggc aca cat tca tct ttg            419
Asn Xaa Tyr Xaa Xaa Trp Gly Gln Gly Thr His Ser Ser Leu
            40              45
targaaggtc tggcctgggg tcrggtgaag gagggcccag gtcagttctg gggtcccagt   479
gacctgcttt gccattctcc tggtgccgct gctgctcct gtttctggag ctggatgttc    539
cccacctggc agttgagctg cctgagccaa tgtgtctgtc tttggtaact gagtgaacca   599
taataaaggg gaacatttgg ccctgtgaaa aaaaaaaa                          637
```

```
<210> 260
<211> 1315
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 123..302

<221> sig_peptide
<222> 123..176
<223> Von Heijne matrix
      score 4.30000019073486
      seq WTCLKSFPSPTSS/HA

<221> polyA_signal
<222> 1279..1284

<221> polyA_site
<222> 1301..1312
```

```
<400> 260
aagagcatcc tgcgccccgg cgcggggccc tgcggtagcc tcaggcccct cccctggacc      60
cgccgcagag ccagtgcaga atacagaaac tgcagccatg accacgcacg tcaccctgga     120
ag atg ccc tgt cca acg tgg acc tgc ttg aag agc ttc ccc tcc ccg        167
   Met Pro Cys Pro Thr Trp Thr Cys Leu Lys Ser Phe Pro Ser Pro
          -15           -10            -5
acc agc agc cat gca tcg agc ctc cac ctt cct cca tca tgt acc agg        215
Thr Ser Ser His Ala Ser Ser Leu His Leu Pro Pro Ser Cys Thr Arg
         1            5                10
cta act ttg aca caa act ttg agg aca gga atg cat ttg tca cgg gca        263
Leu Thr Leu Thr Gln Thr Leu Arg Thr Gly Met His Leu Ser Arg Ala
    15           20               25
ttg caa ggt aca ttg acc agg cta cag tcc act cca gca tgaatgarat        312
Leu Gln Gly Thr Leu Thr Arg Leu Gln Ser Thr Pro Ala
30              35               40
gctggaggaa ggacatgakt atgcggtcat gctgtacacc tggcgcagct gttcccgggc     372
cattccccag gtgaaatgca acragcagcc caaccgakta raratctatg araaracagt     432
araggtgctg gagccggagg tcaccaagct catgaagttc atgtattttc arcgcaaggc     492
catcgagcgg ttctgcascg aggtgaagcg gctgtgccat gccgagcgca ggaaggactt     552
tgtctctgag gcctacctcc tgacccttgg caagttcatc aacatgtttg ctgtcctgga     612
tgagctaaag aacatgaast gcagcgtcaa raatgaccac tctgcctaca agagggcagc     672
acagttcctg cggaagatgg cagatcccca gtctatccag gagtcgcaga acctttccat     732
gttcctggcc aaccacaaca ggatcaccca gtgtctccac cagcaacttg aagtgatccc     792
aggctatgag gagctgctgg ctgacattgt caacatctgt gtggattact acgagaacaa     852
gatgtacctg actcccagtg agaaacatat gctcctcaag gtaaaactcc cctgaggccg     912
cacccatgga gcctgggctt accctctcac cttcttctta ttaaaaatcc gttttaaaaa     972
acaatgtttc ttttttctta aacattgata cagatcttac ggcacataat ggtttgtaac    1032
ctgttccttt cctgtaatat aaatataccgt agtcaccttt ccagatgtca ttaaggctat    1092
ttctacaatg ttatgtgtaa tgactgccaa gtattctgtt gtattggaac attgtcatgt    1152
aacatatccc ctgtggttgg atatttgcta aacttcattg aacacccttg tagcagtttt    1212
tgtgcacatc tttttgtcaa ggcaaacttc ctagaagaga aattgctggc tcaaagggaa    1272
aaacagaata aatcgttttt tttatttcaa aaaaaaaaaa ccc                      1315
```

```
<210> 261
<211> 1035
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 98..673

<221> sig_peptide
<222> 98..376
<223> Von Heijne matrix
     score 5.59999990463257
     seq VLLLRQLFAQAEK/WY

<221> polyA_site
<222> 1025..1035

<400> 261
aattttcygt ggtccaacta ccctcggcga teccaggctt ggcggggcac cgcctggcct        60
ctcccgttcc tttaggctgc cgccgctgcc tgccgcc atg gca gag ttg ggc cta        115
                                           Met Ala Glu Leu Gly Leu
                                                             -90

aat gag cac cat caa aat gaa gtt att aat tat atg cgt ttt gct cgt        163
Asn Glu His His Gln Asn Glu Val Ile Asn Tyr Met Arg Phe Ala Arg
        -85                 -80                 -75

tca aag aga ggc ttg aga ctc aaa act gta gat tcc tgc ttc caa gac        211
Ser Lys Arg Gly Leu Arg Leu Lys Thr Val Asp Ser Cys Phe Gln Asp
    -70                 -65                  -60

ctc aag gag agc agg ctg gtg gag gac acc ttc acc ata gat gaa gtc        259
Leu Lys Glu Ser Arg Leu Val Glu Asp Thr Phe Thr Ile Asp Glu Val
-55                 -50                  -45                 -40

tct gaa gtc ctc aat gga tta caa gct gtg gtt cat agt gag gtg gaa        307
Ser Glu Val Leu Asn Gly Leu Gln Ala Val Val His Ser Glu Val Glu
                -35                  -30                 -25

tct gag ctc atc aac act gcc tat acc aat gtg tta ctt ctg cga cag        355
Ser Glu Leu Ile Asn Thr Ala Tyr Thr Asn Val Leu Leu Leu Arg Gln
            -20                 -15                 -10

ctg ttt gca caa gct gag aag tgg tat ctt aag cta cag aca gac atc        403
Leu Phe Ala Gln Ala Glu Lys Trp Tyr Leu Lys Leu Gln Thr Asp Ile
        -5                  1                   5

tct gaa ctt gaa aac cga gaa tta tta gaa caa ktt gca gaa ttt gaa        451
Ser Glu Leu Glu Asn Arg Glu Leu Leu Glu Gln Xaa Ala Glu Phe Glu
10                  15                  20                  25

aaa gca rav att aca tct tca aac aaa aag ccc atc tta dat gtc aca        499
Lys Ala Xaa Ile Thr Ser Ser Asn Lys Lys Pro Ile Leu Xaa Val Thr
            30                  35                  40

aas cca aaa ctt gct cca ctt aat gaa ggt gga aca gca aaa ctc cta        547
Xaa Pro Lys Leu Ala Pro Leu Asn Glu Gly Gly Thr Ala Lys Leu Leu
        45                  50                  55

aac aag gta ata tgt att att ttg aga aac gga aag tct ctc att ctg        595
Asn Lys Val Ile Cys Ile Ile Leu Arg Asn Gly Lys Ser Leu Ile Leu
        60                  65                  70

tcc tgt cat tgc cta ggg tgg aga aac aaa agt gga agg ttt gtt tca        643
Ser Cys His Cys Leu Gly Trp Arg Asn Lys Ser Gly Arg Phe Val Ser
    75                  80                  85

ggt cct ctg agg ata att agt cca ttg cag tagttttact tgatggtacc        693
Gly Pro Leu Arg Ile Ile Ser Pro Leu Gln
90                  95

ccatgggcca gaagagggca tacttaacct tctagagagc ctgaagtagc tcctgatcac       753
acctttcaa ggtaaagtga agagcatgaa attttggaca gcgtttattg atggacattt        813
aaagttgtg atctgcggta acaaggagaa gggttttttaa gtttataaaa attatttatc       873
aattagccgg gtgtggtggt acgtgcctat agtcagagct actcgggagg ctgaggcagg       933
agaattgctt gaacccggga ggtggaggtt gcagtgagct gagatcacgc cactgcactc       993
tagcctgggc gacagagcga gactccatct caaaaaaaaa aa                        1035
```

<210> 262
<211> 696
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 17..463

<221> sig_peptide
<222> 17..232
<223> Von Heijne matrix
       score 3.79999995231628
       seq LMGLALAVYKCQS/MG

<221> polyA_signal
<222> 657..662

<221> polyA_site
<222> 684..696

<400> 262
```
actcaaacag attccc atg aat ctc ttc atc atg tac atg gca ggc aat act      52
               Met Asn Leu Phe Ile Met Tyr Met Ala Gly Asn Thr
                       -70                     -65
atc tcc atc ttc cct act atg atg gtg tgt atg atg gcc tgg cga ccc        100
Ile Ser Ile Phe Pro Thr Met Met Val Cys Met Met Ala Trp Arg Pro
-60                 -55                     -50                 -45
att cag gca ctt atg gcc att tca gcc act ttc aag atg tta gaa agt       148
Ile Gln Ala Leu Met Ala Ile Ser Ala Thr Phe Lys Met Leu Glu Ser
               -40                     -35                     -30
tca agc cag aag ttt ctt cag ggt ttg gtc tat ctc att ggg aac ctg       196
Ser Ser Gln Lys Phe Leu Gln Gly Leu Val Tyr Leu Ile Gly Asn Leu
           -25                     -20                     -15
atg ggt ttg gca ttg gct gtt tac aag tgc cag tcc atg gga ctg tta       244
Met Gly Leu Ala Leu Ala Val Tyr Lys Cys Gln Ser Met Gly Leu Leu
           -10                     -5                      1
cct aca cat gca tcg gat tgg tta gcc ttc att gag ccc cct gag aga       292
Pro Thr His Ala Ser Asp Trp Leu Ala Phe Ile Glu Pro Pro Glu Arg
5                   10                      15                  20
atg gag tca gtg gtg gag gac tgc ttt tgt gaa cat gag aaa gca gcg       340
Met Glu Ser Val Val Glu Asp Cys Phe Cys Glu His Glu Lys Ala Ala
                   25                      30                  35
cct ggt ccc tat gta ttt ggg tct tat tta cat cct tct tta agc cca       388
Pro Gly Pro Tyr Val Phe Gly Ser Tyr Leu His Pro Ser Leu Ser Pro
               40                      45                  50
gtg gct cct cag cat act ctt aaa cta atc act tat gtt aaa aaa aac       436
Val Ala Pro Gln His Thr Leu Lys Leu Ile Thr Tyr Val Lys Lys Asn
           55                      60                  65
caa aaa act ctt ttc tcc atg gtg ggg tgacaggtcc taaaaggaca            483
Gln Lys Thr Leu Phe Ser Met Val Gly
           70                      75
atgtgcatat tacgacaaac acaaaaaaac tataccataa cccagggctg aaaataatgt    543
aaaaaacttt attttgtttt ccagtacaga gcaaacaac aacaaaaaaa cataactatg    603
taaacaaaaa aataactgct gctaaatcaa aaactgttgc agcatctcct ttcaataaat    663
taaatggttg araacaatgc aaaaaaaaaa aaa                                 696
```

<210> 263
<211> 868

<210> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 263..481

<221> sig_peptide
<222> 263..322
<223> Von Heijne matrix
      score 11.1999998092651
      seq ILVVLMGLPLAQA/LD

<221> polyA_site
<222> 858..868

<400> 263
aagacacgcc tacgattaga ctcaggcagg cacctaccgg cgagcggccg cɪ ʌgtgactc   60
ccaggcgcgg cggtacctca cggtggtgaa ggtcacaggg ttgcagcact cccagtagac   120
caggagctcc gggaggcagg gccggcccca cgtcctctgc gcaccaccct gagttggatc   180
ctctgtgcgc cacccctgag ttggatccag ggctagctgc tgttgacctc cccactccca   240
cgctgccctc ctgcctgcag cc atg acg ccc ctg ctc acc ctg atc ctg gtg   292
                          Met Thr Pro Leu Leu Thr Leu Ile Leu Val
                          -20   -15
gtc ctc atg ggc tta cct ctg gcc cag gcc ttg gac tgc cac gtg tgt   340
Val Leu Met Gly Leu Pro Leu Ala Gln Ala Leu Asp Cys His Val Cys
-10   -5   1   5
gcc tac aac gga gac aac tgc ttc aac ccc atg cgc tgc ccg gct atg   388
Ala Tyr Asn Gly Asp Asn Cys Phe Asn Pro Met Arg Cys Pro Ala Met
              10   15   20
gtt gcc tac tgc atg acc acg cgc acc tac tac acc ccc acc agg atg   436
Val Ala Tyr Cys Met Thr Thr Arg Thr Tyr Tyr Thr Pro Thr Arg Met
              25   30   35
aag gtc agt aag tcc tgc gtg ccc cgc tgc ttc gar nac tgt gta   481
Lys Val Ser Lys Ser Cys Val Pro Arg Cys Phe Glu Xaa Cys Val
              40   45   50
tgatggctac tccaagcacg cgtccaccac ctcctgctgc cagtacgacc tctgcaacgg   541
caccggcctt gccaccccgg ccaccctggc cctggccccc atcctcctgg ccaccctctg   601
gggtctcctc taaagccccc gaggcagacc cactcaagaa caaagctctc gagacacact   661
gctayaccct ckcacccakc tcaccctgcc tcaccctcca cactccctgc gacctcctca   721
gccatgccca gggtcaggac tgtgggcaag aagacacccg acctccccca accaccacac   781
gacctcactt cgaggccttg acctttcgat gctgtgtggg atcccaaaag tgtccggctt   841
tgatgggctg atcagcaaaa aaaaaaa   868

<210> 264
<211> 775
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 42..299

<221> sig_peptide
<222> 42..101
<223> Von Heijne matrix
      score 5.40000009536743
      seq WFVHSSALGLVLA/PP

<221> polyA_site

<222> 762..775

<400> 264
```
aacgatacaa atggtaggcc ttcatgtgag ccagtdacta c atg aat ctt cat ttc    56
                                              Met Asn Leu His Phe
                                                             -20
cca cag tgg ttt gtt cat tca tca gcg tta ggc ttg gtc ctg gct cca   104
Pro Gln Trp Phe Val His Ser Ser Ala Leu Gly Leu Val Leu Ala Pro
-15             -10             -5                          1
cct ttc tcc tct ccg ggc act gac ccc acc ttt ccg tgt att tac tgt   152
Pro Phe Ser Ser Pro Gly Thr Asp Pro Thr Phe Pro Cys Ile Tyr Cys
             5               10              15
agg cta tta aat atg atc atg acc cgc ctt gca ttt tca ttc atc acc   200
Arg Leu Leu Asn Met Ile Met Thr Arg Leu Ala Phe Ser Phe Ile Thr
        20              25              30
tgt tta tgc cca aat tta aag gaa gtt tgt ctc att ttg cca gaa aaa   248
Cys Leu Cys Pro Asn Leu Lys Glu Val Cys Leu Ile Leu Pro Glu Lys
    35              40              45
aat tgt aat agt cga cac gct gga ttt gta ggg cca sca aaa ttg cgg   296
Asn Cys Asn Ser Arg His Ala Gly Phe Val Gly Pro Xaa Lys Leu Arg
50              55              60              65
cag tgaaactwkk ttcwcttcta aagcccttca tttcccacaa ggttaagctc        349
Gln
tcgaaacccc atttgatcct tggttcctat ttcgatcctc ctttggaatc tgaaaatcgg   409
tctccatgtt gtatgcaaat taaaakttgc cttgtttgtt actcttccaa cacagggtat   469
cagggaraaa gaggccttat ctgttcctcc atcccccctg ttttgacaga ctgctaagaa   529
ttcctcagga cttcctttgg ttggggattt tactttccca aaagtctgat ctgatttctt   589
tcaggggtag acaagcttgt cctagtgctc tgcttcaggt cttatcagaa gaaacccagg   649
aatagaaaag gtagatgcct tgactttgt ccctgttgtg gggactaaag tgttttttgc   709
cagaattgtc aaaagctccg gttcaaactc tgtagagttt catggaaaaa caaaacaaaa   769
aaaaaa                                                              775
```

<210> 265
<211> 1075
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 198..431

<221> sig_peptide
<222> 198..260
<223> Von Heijne matrix
       score 6.90000009536743
       seq LLACGSLLPGLWQ/HL

<221> polyA_site
<222> 1064..1074

<400> 265
```
atatatttct gaggcagtac ccatctcact tgtaaactta aaagacaccg cagagatttg    60
agggactcag aagtcaaata gagtaggtta aaaacctctt attttttcaaa ttaattgttt   120
taagaaacaa gcatacctgt gtaagtgaaa tatcttaatt tgtgttgaat caagttagga   180
gacagagatt ctcatga atg tgt cct gtg ttc tca aag cag ctg cta gcc       230
                    Met Cys Pro Val Phe Ser Lys Gln Leu Leu Ala
                    -20                     -15
tgt ggg tct ctc cta cct ggg tta tgg cag cac ctc aca gcc aat cac     278
Cys Gly Ser Leu Leu Pro Gly Leu Trp Gln His Leu Thr Ala Asn His
-10             -5              1               5
```

```
tgg cct cca ttc tcc sct ttc ctc tgt aca gtt tgc tct ggt tcc tca     326
Trp Pro Pro Phe Ser Xaa Phe Leu Cys Thr Val Cys Ser Gly Ser Ser
            10                  15                  20
gag cag att tcc gag tat act gct tca gcc acg ccc cca ctg tgc cgt     374
Glu Gln Ile Ser Glu Tyr Thr Ala Ser Ala Thr Pro Pro Leu Cys Arg
        25                  30                  35
tcc ctg aac caa gag cca ttc gty tca aga gcc att cgt cca aag tac     422
Ser Leu Asn Gln Glu Pro Phe Val Ser Arg Ala Ile Arg Pro Lys Tyr
        40                  45                  50
tct atc acc tagccattgt akccatacca agccgggctt cctacttccc            471
Ser Ile Thr
55
tctgctcccc ttggtttcct cctgtraart aaatctcact gacccttgat gcasctccaa    531
gcatatataa tatatatata ataaaaccat abtctaaaaa attcaaacca ggawaaataa    591
asccaraaat ttgtatggga aaaatctgca caaatttatt tggccagcat ggttatcatg    651
gctctattga atttatcctt gaccgtcttt aaagccaaag caaacgggat aaagtgatca    711
actacttacc tctcaataac aaaaargaag caggaggcaa aatctctcaw taatttcata    771
aaaacaattc ttakctgggc gcggtggctc wcacctgtar tcccaacact ttjggaggcc    831
saggtgggcg gatcatgagg tcgggagatc aamaccatcc tggctaacat ggtgaaaccc    891
catctctact aaaattacaa aaaattrgct gggcgaggtg gcgggcacct gtggtcccag    951
ctactcggga ggctgaggca agagaatggt gtgaacccca gggggcggag cctgcagtga   1011
gctgagatcg caccactgca ctccagcctg ggcgacagtg agactccgtc tcaaaaaaaa   1071
aaah                                                               1075
```

```
<210> 266
<211> 981
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 279..473

<221> sig_peptide
<222> 279..362
<223> Von Heijne matrix
      score 4.40000009536743
      seq SCFLVALIIWCYL/RE

<221> polyA_signal
<222> 944..949

<221> polyA_site
<222> 970..981

<400> 266
agaatcgtgt cttgtgtgcc ccggcggccg ggtgagctcc tcaaggtctc ggagggccga    60
gggcagacac cggcgggcgg gcggasgctt actgctctct ctcttccagg gccgtccggg   120
cgctgaggct cataggctgg gcttcccgaa gccttcatcc gttgcccggt tcccgggatc   180
gggcccaccc tgccgccgag gaagaggacg accctgaccg ccccattgag ttttcctcca   240
gcaaagccaa ccctcaccgc tggtcggtgg gccatacc atg gga aag gga cat cag   296
                                         Met Gly Lys Gly His Gln
                                                            -25
cgg ccc tgg tgg aag gtg ctg ccc ctc agc tgc ttc ctc gtg gcg ctg     344
Arg Pro Trp Trp Lys Val Leu Pro Leu Ser Cys Phe Leu Val Ala Leu
        -20                 -15                 -10
atc atc tgg tgc tac ctg agg gag gag agc gag gcg gac cag tgg ttg     392
Ile Ile Trp Cys Tyr Leu Arg Glu Glu Ser Glu Ala Asp Gln Trp Leu
        -5                  1               5                   10
aga cag gtg tgg gga gag gtg cca gag ccc agt gat cgt tct gag gag     440
```

```
Arg Gln Val Trp Gly Glu Val Pro Glu Pro Ser Asp Arg Ser Glu Glu
              15                  20                  25
cct gag act cca gct gcc tac aga gcg aga act tgacggggtg cccgctgggg   493
Pro Glu Thr Pro Ala Ala Tyr Arg Ala Arg Thr
              30                  35
ctggcaggaa gggagccgac asccgccctt cggatttgat ktcacgtttg cccgtgactg   553
tcctggctat gcktgcgtcc tcagcactra argacttggc tggtggatgg ggcacttggc   613
tatgctgatt cgcgtgaagg cggavcaaaa tctcagcaaa tcggaaactg ctcctcscct   673
ggctcttgat ktccaaggat tccatcggca aaacttctca ratccttggg gaaggtttca   733
gttgcactgt atgctgttgg atttgccaag tctttgtata acataatcat gtttccaaag   793
cacttctggt gacacttgtc atccagtgtt agtttgcagg taatttgctt tctgagatag   853
aatatctggc agaagtgtga aactgtattg catgctgcgg cctgtgcaag gaacacttcc   913
acatgtgagt tttacacaac aacaaatgaa aataaatttt aattttataa tatgggaaaa   973
aaaaaaaa                                                             981
```

<210> 267
<211> 1031
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 12..644

<221> sig_peptide
<222> 12..92
<223> Von Heijne matrix
      score 4
      seq LTFFSGVYGTCIG/AT

<221> polyA_signal
<222> 1002..1007

<221> polyA_site
<222> 1020..1031

<400> 267
```
acaccaagga g atg ctc ctt ctt agt att aca act gct tat aca ggt ctg   50
            Met Leu Leu Leu Ser Ile Thr Thr Ala Tyr Thr Gly Leu
                    -25                  -20                  -15
gaa tta act ttc ttc tct ggt gta tat gga acc tgt att ggt gct aca   98
Glu Leu Thr Phe Phe Ser Gly Val Tyr Gly Thr Cys Ile Gly Ala Thr
            -10                  -5                    1
aat aaa ttt gga gca gaa gag ara agc ctt att gga ctt tct ggc att   146
Asn Lys Phe Gly Ala Glu Glu Xaa Ser Leu Ile Gly Leu Ser Gly Ile
        5                    10                  15
ttc atc ggc att gga gaa att tta ggt gga agc ctc ttc ggc ctg ctg   194
Phe Ile Gly Ile Gly Glu Ile Leu Gly Gly Ser Leu Phe Gly Leu Leu
    20                  25                  30
agc aag aac aat cgt ttt ggt aga aat cca gtt gtg ctg ttg ggc atc   242
Ser Lys Asn Asn Arg Phe Gly Arg Asn Pro Val Val Leu Leu Gly Ile
35                  40                  45                  50
ctg gtg cac ttc ata gct ttt tat cta ata ttt ctc aac atg cct gga   290
Leu Val His Phe Ile Ala Phe Tyr Leu Ile Phe Leu Asn Met Pro Gly
                    55                  60                  65
gat gcc ccg att gct cct gtt aaa gga act gac agc agt gct tac atc   338
Asp Ala Pro Ile Ala Pro Val Lys Gly Thr Asp Ser Ser Ala Tyr Ile
                70                  75                  80
aaa tcc agc aaa raa ttt gcc att ctc tgc akt ttt ctg tkg ggc ctt   386
Lys Ser Ser Lys Xaa Phe Ala Ile Leu Cys Xaa Phe Leu Xaa Gly Leu
```

```
         85              90              95
gga aac agc tgc ttt aat acc cas ctg ctt akt atc tkg ggc ttt ctg    434
Gly Asn Ser Cys Phe Asn Thr Xaa Leu Leu Xaa Ile Xaa Gly Phe Leu
    100             105             110
tat tct gaa rac agc gcc cca kca ttt gcc atc ttc aat ttt gtt cag    482
Tyr Ser Glu Xaa Ser Ala Pro Xaa Phe Ala Ile Phe Asn Phe Val Gln
115             120             125             130
tct att tgc gca gcc gtg gca ttt ttc tac agc aac tac ctt ctc ctt    530
Ser Ile Cys Ala Ala Val Ala Phe Phe Tyr Ser Asn Tyr Leu Leu Leu
            135             140             145
cac tgg caa ctc ctg gtc atg gtk atw ttt ggg ttt ttk gga aca att    578
His Trp Gln Leu Leu Val Met Val Ile Phe Gly Phe Xaa Gly Thr Ile
            150             155             160
tct ttc ttc act gtg gaa tgg gaa sct gcc gcc ttt gta scc cgc ggc    626
Ser Phe Phe Thr Val Glu Trp Glu Xaa Ala Ala Phe Val Xaa Arg Gly
            165             170             175
tct gac tac cga agt atc tgatctggtg tccgtgaggg gacacgtatg          674
Ser Asp Tyr Arg Ser Ile
            180
acctcagaaa cacagctgga cacagagctt ggtggaagaa gtcgcctttg atcttcacta  734
tatattgggt gatgttcagt atggaaaatc aagggattaa gactgttaaa tcagccagag  794
tkggtgttca agtttacaga tatgagttat ttaaagcaag tagaataagg gaaagctgtt  854
ctgtcaactg taattgttca agatgttgt ttttcattc atctatctca attcttataa   914
tcatgttata gaatgtaaat gttttcttct ctctcctgct cttgttggaa gatcctgcct  974
tgatttagaa tactaggcca tatgtcatat aaatattttt ctggaaaaa aaaaaaa     1031
```

<210> 268
<211> 1283
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 91..459

<221> sig_peptide
<222> 91..330
<223> Von Heijne matrix
      score 7.69999980926514
      seq LVLFLSLALLVTP/TS

<221> polyA_site
<222> 1271..1281

<400> 268
```
tattccttgg agttccacga ctgaattaag actgttgtgg grdccataat tttcaaatac   60
ttgccctata ttcgtgttga gggttcacac atg agc aca tgg tat ttg gca ctt  114
                                   Met Ser Thr Trp Tyr Leu Ala Leu
                                       -80             -75
aat aag tcc tat aag aat aaa gac agc gtt acg att tat ctc agc ttg    162
Asn Lys Ser Tyr Lys Asn Lys Asp Ser Val Arg Ile Tyr Leu Ser Leu
    -70             -65             -60
tgc aca gtg agc att aaa ttt aca tac ttt cat gat ata cag act aat    210
Cys Thr Val Ser Ile Lys Phe Thr Tyr Phe His Asp Ile Gln Thr Asn
    -55             -50             -45
tgt ctt aca aca tgg aaa cat tcg aga tgc aga ttt tat tgg gca ttt    258
Cys Leu Thr Thr Trp Lys His Ser Arg Cys Arg Phe Tyr Trp Ala Phe
    -40             -35             -30             -25
ggt ggt tcc att tta cag cac tca gtg gat ccc ctt gtt ttg ttc cta    306
Gly Gly Ser Ile Leu Gln His Ser Val Asp Pro Leu Val Leu Phe Leu
```

```
                -20                 -15                 -10
agc ctg gcc ctg tta gtg aca ccc act tcc acc cct tct gct aar ata    354
Ser Leu Ala Leu Leu Val Thr Pro Thr Ser Thr Pro Ser Ala Lys Ile
            -5                   1                   5
car agc ctt caa att gac ctc cct gga ggc tgg agg ctg gcc act gac    402
Gln Ser Leu Gln Ile Asp Leu Pro Gly Gly Trp Arg Leu Ala Thr Asp
         10                  15                  20
agg atc ttt acc ctc tcc ccc gta ccc atg gac rgc ccc ctc atc ctt    450
Arg Ile Phe Thr Leu Ser Pro Val Pro Met Asp Xaa Pro Leu Ile Leu
25                      30                  35                  40
cat cag ttg taaaggtaga tatttgttcc ttggagtcca acatcatgct            499
His Gln Leu
gttcagaata taatgagatc aatagttgaa aaactagata tacatgccac ccwgacaaag   559
ctattaagtt attaagtgtc agccctggat cttggcttat tgtgaaatgt taattatttt   619
atcactcyat taagaagctg tgggctccat ctcagcattg aaaagggact aatttgctct   679
gttttggaat tgaattagct ttcaggccas cagggcactg tttggtaaat tgcttttttcc  739
agtactagca tgttttctcc ctccatagcc tctgttagct tctgagcttg taacctccag   799
ggaaavatga gaatattcac cctttttaata tgtgtagaga ccatgcaaga ccattgtctt   859
ctaataatta gaaatactta gccagattct ctatagtaaa cccggagatt gggagggctg   919
ctttctactt ggtgcatcct tctgcgcttc taatgatttt taaaaatctg ttaataattg    979
atgttttctg gctgggcaca gtggctcacg cctgtaatcc cagcactttg ggaggccaag   1039
gagggcagat catgaggtca ggagattgar accatcctgg ctaacacggt gaaaccccgt   1099
ctctactaaa aatacaaaar aattakccgg gcatggtagt gggcgcctgt gtacccagct   1159
actggggagg ctgaggcarg araatcgctt gaacctggga ggcggaggtt gcastragct   1219
gagatggtgc caccgcactc tagcctgggt gacagagcga gacttcattt caaaaaaaaa   1279
aamc                                                               1283
```

```
<210> 269
<211> 1777
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 70..327

<221> sig_peptide
<222> 70..147
<223> Von Heijne matrix
      score 9.60000038146973
      seq WLIALASWSWALC/RI

<221> polyA_signal
<222> 1741..1746

<221> polyA_site
<222> 1763..1774

<400> 269
agcccggttt cgtgcccgcg gccgactgcg casctgtccg cgagtctgag atacttacag    60
agagctaca atg gaa aag tcc tgg atg ctg tgg aac ttt gtt gaa aga tgg   111
           Met Glu Lys Ser Trp Met Leu Trp Asn Phe Val Glu Arg Trp
               -25                 -20                 -15
cta ata gcc ttg gct tca tgg tct tgg gct ctc tgc cgt att tct ctt    159
Leu Ile Ala Leu Ala Ser Trp Ser Trp Ala Leu Cys Arg Ile Ser Leu
        -10                  -5                   1
tta cct tta ata gtg act ttt cat ctg tat gga ggc att atc tta ctt    207
Leu Pro Leu Ile Val Thr Phe His Leu Tyr Gly Gly Ile Ile Leu Leu
5                   10                  15                  20
ttg tta ata ttc ata tca atw kca ggt att ctg tat aaa ttc cas gat    255
```

```
Leu Leu Ile Phe Ile Ser Ile Xaa Gly Ile Leu Tyr Lys Phe Xaa Asp
              25                30                35
gta ttg ctt tat ttt ccw kaa cag yya tcc tct tca cgt ctt tat gat     303
Val Leu Leu Tyr Phe Pro Xaa Gln Xaa Ser Ser Ser Arg Leu Tyr Asp
              40                45                50
tcc cat gcc cac tgg cmt tcg rca taaaaaaatt ttcatcagaa ccaaagatgg    357
Ser His Ala His Trp Xaa Ser Xaa
              55                60
aatacgtctg aatcttattt tgatacgata cactggagac aattcaccct attccccaac   417
tataatttat tttcatggga atgcaggcaa cataggtcac aggttggcca aatgcattac   477
ttatgttggt taacctcaaa gttaaccttt tgctggttga ttatcgagga tatggaaaaa   537
gtgaaggaga agcaagtgaa gaaggactct acttagattc tgaagctgtg ttagactacg   597
tgatgactag acctgacctt gataaaacaa aaatttttct ttttggccgt tccttgggtg   657
garcagtggc tattcatttg gcttctgaaa attcacatag gatttcagcc attatggtgg   717
agaacacatt tttaagcata ccacatatgg cagcacttt attttcattc tttccgatgc    777
gttaccttcc tttatggtgc tacaaaaata aattttttgtc ctacagaaaa atctctcagt   837
gtagaatgcc ttcacttttc atctctggac tctcagatca attaattcca ccagtaatga   897
tgaaacaact ttatgaactc tccccatctc ggactaagan attagccatt tttccagatg   957
ggactcacaa tgacacatgg cagtgccaag gctatttcac tgcacttgaa cagttcatca   1017
aagaagtcgt aaagagccat tctcctgaag aaatggcaaa aacttcatct aatgtaacaa   1077
ttatataatg tttccctttt tgattattgc attgcatttt aatttgtgca gaatgataaa   1137
gaatgttcct tttagaagtg tgttatgtct gtacctgtct gaagagtgac attaaacttt   1197
gaaaggactt cactgctcct ttacgatatt ccaaatagtt ttttacattg gaaaaactaa   1257
ttcttgggat tctttcatac attttcatca aaactttcag tgtgattatg tattcatatc   1317
ttcagtttaa tatgtcagta taatagatat tgttcaaaag tttcttgttg ctaaagtggt   1377
gtaatctgtt acacagatga atagctagat gtggaaagag atatgtaaac aagaaacctt   1437
tgggtattgt ttcttaagta aatattggga caatcatggt aagcaaactt agttctgtaa   1497
ctgcattttt caccttaaaa gttaaatgaa atgcatgatg gtattttatt ccttgaatta   1557
tgcaatgcaa cattttacat gtaaatagca ctggtcatat actgatgtat atggttatct   1617
gggttatatc tattttttatg taaactctat ttttgttttt ggcaagaagt gaaattgaga   1677
cttatgtgca ggttgccatt gaattttgct ctggtgaatg ctgagatcca gctttttctt   1737
acaaataaat gggaccctgt tttccaaaaa aaaaaaamcm                          1777
```

```
<210> 270
<211> 970
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 12..497

<221> sig_peptide
<222> 12..104
<223> Von Heijne matrix
      score 5.5
      seq LVGVLWFVSVTTG/PW

<221> polyA_signal
<222> 935..940

<221> polyA_site
<222> 955..967

<400> 270
aggtctccaa g atg gcg gcc gcc tgg ccg tct ggt ccg kct gct ccg gag     50
            Met Ala Ala Ala Trp Pro Ser Gly Pro Xaa Ala Ala Pro Glu
                -30                -25                -20
gcc gtg acg gcc aga ctc gtt ggt gtc ctg tgg ttc gtc tca gtc act      98
Ala Val Thr Ala Arg Leu Val Gly Val Leu Trp Phe Val Ser Val Thr
```

```
                -15                    -10                    -5
aca gga ccc tgg ggg gct gtt gcc acc tcc gcc ggg ggc gag gag tcg       146
Thr Gly Pro Trp Gly Ala Val Ala Thr Ser Ala Gly Gly Glu Glu Ser
                 1               5                       10
ctt aag tgc gag gac ctc aaa gtg gga caa tat att tgt aaa gat cca       194
Leu Lys Cys Glu Asp Leu Lys Val Gly Gln Tyr Ile Cys Lys Asp Pro
15                  20                      25                      30
aaa ata aat gac gct acg caa gaa cca gtt aac tgt aca aac tac aca       242
Lys Ile Asn Asp Ala Thr Gln Glu Pro Val Asn Cys Thr Asn Tyr Thr
                 35                      40                      45
gct cat gtt tcc tgt ttt cca gca ccc aac ata act tgt aag gat tcc       290
Ala His Val Ser Cys Phe Pro Ala Pro Asn Ile Thr Cys Lys Asp Ser
                 50                      55                      60
agt ggc aat gaa aca cat ttt act ggg aac gaa gtt ggt ttt ttc aag       338
Ser Gly Asn Glu Thr His Phe Thr Gly Asn Glu Val Gly Phe Phe Lys
                 65                      70                      75
ccc ata tct tgc cga aat gta aat ggc tat tcc tac aat gag cag tcg       386
Pro Ile Ser Cys Arg Asn Val Asn Gly Tyr Ser Tyr Asn Glu Gln Ser
                 80                      85                      90
cat gtc tct ttt tct tgg atg gtt ggg agc aga tcg att tta cct tgg       434
His Val Ser Phe Ser Trp Met Val Gly Ser Arg Ser Ile Leu Pro Trp
95                  100                     105                     110
ata ccc tgc ttt ggg ttt gtt aaa btt tyg cac tgt agg gtt tkg tgg       482
Ile Pro Cys Phe Gly Phe Val Lys Xaa Xaa His Cys Arg Val Xaa Trp
                 115                     120                     125
aat tgg gag cct aat tgatttcaty cttatttcaa tgcagattgt tggaccttca       537
Asn Trp Glu Pro Asn
                 130
aatggaagta gttacattat agattactat ggaaccagac ttacaagact gagtattact      597
aatgaaacat ttagaaaaac gcaattatat ccataaatat tttttaaaag aaacagattt      657
gagcctcctt gattttaata gagaacttct agtgtatgga tttaaagatt tctctttttc      717
attcatatac cattttatga gttctgtata attttttgtg gtttttgttt tgttgagtta      777
aagtatatta ttgtgagatt tatttaatag gacttccttt gaaagctgta taatagtgtt      837
tctcgggctt ctgtctctat gagagatagc ttattactct gatactcttt aatcttttac      897
aaaggcaagt tgccacttgt cattttgtt tctgaaaaat aaaagtataa cttattcaca        957
aaaaaaaaaa mms                                                          970
```

```
<210> 271
<211> 645
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 90..383

<221> sig_peptide
<222> 90..200
<223> Von Heijne matrix
      score 4.90000009536743
      seq MLIMLGIFFNVHS/AV

<221> polyA_signal
<222> 609..614

<221> polyA_site
<222> 632..643

<400> 271
atctctgccc ccctgcgagg gcatcctggg ctttctccca ccgctttccg agcccgcttg       60
```

```
cacctcggcg atccccgact cccttcttt atg gcg tcg ctc ctg tgc tgt ggg      113
                                   Met Ala Ser Leu Leu Cys Cys Gly
                                      -35                     -30
ccg aag ctg gcc gcc tgc ggc atc gtc ctc agc gcc tgg gga gtg atc      161
Pro Lys Leu Ala Ala Cys Gly Ile Val Leu Ser Ala Trp Gly Val Ile
           -25                 -20                 -15
atg ttg ata atg ctc gga ata ttt ttc aat gtc cat tcc gct gtg ttg      209
Met Leu Ile Met Leu Gly Ile Phe Phe Asn Val His Ser Ala Val Leu
           -10                 -5                  1
att gag gac gtt ccc ttc acg gag aaa gat ttt gag aac ggc ccc car      257
Ile Glu Asp Val Pro Phe Thr Glu Lys Asp Phe Glu Asn Gly Pro Gln
    5                   10                  15
aac ata tac aac ctt tac rag caa ktc agc tac aac tgt ttc atc gct      305
Asn Ile Tyr Asn Leu Tyr Xaa Gln Xaa Ser Tyr Asn Cys Phe Ile Ala
20                  25                  30                  35
gca ggc ctt tac ctc ctc ctc gga ggc ttc tct ttc tgc caa ktt cgg      353
Ala Gly Leu Tyr Leu Leu Leu Gly Gly Phe Ser Phe Cys Gln Xaa Arg
               40                  45                  50
ctc aat aag cgc aag gaa tac atg gtg cgc tagggccccg gcgcgtttcc        403
Leu Asn Lys Arg Lys Glu Tyr Met Val Arg
               55                  60
ccgctccagc ccctcctcta tttaaaract ccctgcaccg tktcacccag gtcgcgtccc    463
acccttgccg gcgccctctg tgggactggg tttcccgggc rararactga atcccttctc    523
ccatctctgg catccggccc ccgtggarar ggctgaggct gggggggctgt tccgtctctc   583
caccccttcgc tgtgtcccgt atctcaataa agagaatctg ctctcttcaa aaaaaaaaaa   643
my                                                                   645
```

```
<210> 272
<211> 773
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 332..541

<221> sig_peptide
<222> 332..376
<223> Von Heijne matrix
      score 3.59999990463257
      seq FLPCCLLWSVFNP/ES

<221> polyA_signal
<222> 739..744

<221> polyA_site
<222> 761..773

<400> 272
aaaacaattc atgcctttca tagtttatta ttattaaagt ctaaacaaaa ttgcaatttc     60
ttaggtaacc ttatatttac aataaatgaa gattaccctc aaatgctaga agctgtctag    120
gtccgtccgg tgtgtcagat tttcctcaga ttagatgtgc caataaccaa gtttattcag    180
taaacaactt gtacttgttt catctggttt tattactctc acccataaac agtaatgact    240
ctctgaccct ctggaaatat gtaatgcttc caatcttgct ttgtgtatct catttaattt    300
gttataaggt agtactgatt ttagcatatt a atg cga ttt ctt cct tgt tgt      352
                                   Met Arg Phe Leu Pro Cys Cys
                                      -15                 -10
ttg ctt tgg tct gtg ttc aat cca gag agc tta aat tgt cat tat ttt      400
Leu Leu Trp Ser Val Phe Asn Pro Glu Ser Leu Asn Cys His Tyr Phe
       -5                  1                   5
```

```
ghk ndd gaa amc tgt att ttt gyt agt tta caa tat tat gaa att tca        448
Xaa Xaa Glu Xaa Cys Ile Phe Xaa Ser Leu Gln Tyr Tyr Glu Ile Ser
        10                  15                  20
ctt cag gag aaa ctg ctg ggc ttc ctg tgg ctt tgt ttt ctt agt tac        496
Leu Gln Glu Lys Leu Leu Gly Phe Leu Trp Leu Cys Phe Leu Ser Tyr
25                  30                  35                  40
ttt ttc cgt gcc gtg tat ttt tta att gat ttt tct tct ttt act          541
Phe Phe Arg Ala Val Tyr Phe Leu Ile Asp Phe Ser Ser Phe Thr
                    45                  50                  55
tgaaaagaaa gtgttttatt ttcaaatctg gtccatattt acattctagt tcagagccaa       601
gccttaaact gtacagaatt tccactgtaa ttaaaactat ttagtgttag ttataaatag       661
ccttcaaaaa gagagattct ccattacacg atcacctgca tcacagccca tggtgaatgt       721
atgtttctgc atagcgaaat aaaaatggca aatgcactga aaaaaaaaa aa                773
```

```
<210> 273
<211> 566
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 43..222

<221> sig_peptide
<222> 43..177
<223> Von Heijne matrix
      score 4
      seq ENFLSLLSKSCSA/DP

<221> polyA_signal
<222> 530..535

<221> polyA_site
<222> 555..566

<400> 273
aacgagtgga ggtgtggcta gtggctgtga tgagataaat cc atg cat agc ctt          54
                                               Met His Ser Leu
                                                -45
ttc att gcg agc ttg aaa gtt ctt ttc tat tac agt ttt agc ttt agg        102
Phe Ile Ala Ser Leu Lys Val Leu Phe Tyr Tyr Ser Phe Ser Phe Arg
    -40                 -35                 -30
ttt aat tgg ttc gac tgc ctt ctc cac aat ttg ggc gag aat ttc ctt        150
Phe Asn Trp Phe Asp Cys Leu Leu His Asn Leu Gly Glu Asn Phe Leu
-25                 -20                 -15                 -10
agc ctt ctc agc aaa agt tgt tct gcg gac ccg tct ggg tca act ttc        198
Ser Leu Leu Ser Lys Ser Cys Ser Ala Asp Pro Ser Gly Ser Thr Phe
                    -5                  1                   5
atg agg gac att gag aca aac aaa tgaaatatgg gttaaagtac tctgagcagc        252
Met Arg Asp Ile Glu Thr Asn Lys
        10                  15
tacaaaaaga araccagtct atcctgctgg agacagtggc cacgtgaara aagagctctt        312
gcagtatgaa agaccacatg gaaagagagg ccacatggaa ccaacagtca gcatcttggt        372
ttcggacacg tgaaraaatt catctcarac tgtgtatcct aaatcaggca cttgctgaat        432
ctaactacat gagtgagacc agttgacaac acatggagca racatgagct gttctcagtg        492
artcctacac aaattcctga ctcacaacac tgtgagcaat aaaatggttg ttattttaag        552
ccaaaaaaaa aaaa                                                          566
```

<210> 274
<211> 455
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 115..231

<221> sig_peptide
<222> 115..180
<223> Von Heijne matrix
      score 5
      seq HLFVTWSSQRALS/HP

<221> polyA_signal
<222> 419..424

<221> polyA_site
<222> 445..455

<400> 274
```
aacctgccag tkatgcaaat gccaaaatgt gggtcatcat atagtatatt tgaaaccttt      60
ctgaacatgt acaccaccca atgctagagg ctgacttgga aaccggtggg tgca atg      117
                                                              Met
ccc gag gct gtg gaa caa tca gcc cat ctc ttt gtg acc tgg agc agt      165
Pro Glu Ala Val Glu Gln Ser Ala His Leu Phe Val Thr Trp Ser Ser
    -20             -15                 -10
cag agg gcc ctc agt cac ccc gcc cca ttc ctc acc ara raa aar aat      213
Gln Arg Ala Leu Ser His Pro Ala Pro Phe Leu Thr Xaa Xaa Lys Asn
-5              1               5                   10
cca ttt cta tgg aag ctc tgacgtaact tcagtgtttt ctacaatact            261
Pro Phe Leu Trp Lys Leu
                15
cctcctgccc cgccccatta aaacagttct tttgttaaaa aatavcctaa tggtccaact      321
ttgctgtctg ttcttccaaa tgtttataat acacattatt tataaatatg tctgtttggg      381
aagctaagaa caagctagtt tttacaacac aaatggaaat aaatgcaatt attataaaaa      441
tycaaaaaaa aaaa                                                        455
```

<210> 275
<211> 673
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 232..384

<221> sig_peptide
<222> 232..300
<223> Von Heijne matrix
      score 3.70000004768372
      seq FFLCAAFPLGAGV/KM

<221> polyA_signal
<222> 650..655

<221> polyA_site
<222> 662..673

```
<400> 275
atttggcttg cagactgcct tctatcccag aacagctgag aaatctatga agctgagatt      60
ctgaaggacc cagcttaggt tcttccactt aggcctcaat tcccttcctt ttccaggggc     120
agccttagtt tcccatggcc ctgaaacaca cacatttccc ccttcctttc ccagaagcca     180
ctggcccccc atagcaccca gtgcatcctt tttacaagtg gaagaactag g atg gct     237
                                                         Met Ala

ttc caa agt ctt cta gaa atg aag ttc ttt ctc tgt gca gct ttc ccc     285
Phe Gln Ser Leu Leu Glu Met Lys Phe Phe Leu Cys Ala Ala Phe Pro
    -20              -15               -10

ctt gga gca gga gtg aag atg ttt cat tat ctt ggg cct ggg aaa cca     333
Leu Gly Ala Gly Val Lys Met Phe His Tyr Leu Gly Pro Gly Lys Pro
-5              1                5                10

ctt cyy cag gct tct ccc tcc ccc cac ccc cat agg amc agg att tgg     381
Leu Xaa Gln Ala Ser Pro Ser Pro His Pro His Arg Xaa Arg Ile Trp
        15               20               25

cct tagcttctgg gcctatcsgc tgccttccct cttyttccta ccacctcttc           434
Pro
tgccttcctt trawctctgt tgggcttggg gatcttagtt ttcttttgtt tatctcccat     494
ctcatttttt tcttctggtc agttttttta aggggggtg ttgtggtttt ttgtttttgt     554
tttgcttctg aaaaarcatt tgcctttctt cctctcccaa cataacaatc gtggtaacag     614
aatgcgactg ctgatttacc gatgtattta atgtaagtaa aaaaaggaaa aaaaraaaa      673


<210> 276
<211> 639
<212> DNA
<213> Homo sapiens.

<220>
<221> CDS
<222> 143..427

<221> sig_peptide
<222> 143..286
<223> Von Heijne matrix
      score 7.5
      seq FVILLLFIFTVVS/LV

<221> polyA_signal
<222> 606..611

<221> polyA_site
<222> 628..639

<400> 276
aatcgcttca gcagcatcct ctcagacaag agccactatt tctgattcag atcacctgtc      60
atcgaagttt aaagaagggg aaacaggaga cagaaataca ctgaaccaaa agattcaaa     120
agagcaagtg gaatctctaa ga atg gct tcc agc cac tgg aat gaa acc act     172
                         Met Ala Ser Ser His Trp Asn Glu Thr Thr
                                               -45               -40

acc tct gtt tat cag tac ctt ggt ttt caa gtt caa aaa att tac cct     220
Thr Ser Val Tyr Gln Tyr Leu Gly Phe Gln Val Gln Lys Ile Tyr Pro
        -35               -30                    -25

ttc cat gac aac tgg aac act gcc tgc ttt gtc atc ctg ctt tta ttt     268
Phe His Asp Asn Trp Asn Thr Ala Cys Phe Val Ile Leu Leu Leu Phe
        -20               -15                    -10

ata ttt aca gtg gta tct tta gtg gtg ctg gct ttc ctt tat gaa gtg     316
Ile Phe Thr Val Val Ser Leu Val Val Leu Ala Phe Leu Tyr Glu Val
-5               1                5                10

ctt gam wgc tgc tgc tgt gta aaa aac aaa acc gtg aaa gac ttg aaa     364
Leu Xaa Xaa Cys Cys Cys Val Lys Asn Lys Thr Val Lys Asp Leu Lys
```

```
                         15                  20                  25
agt gaa ccc aac cct ctt ara akt atg atg gac aac atc aga aaa cgt        412
Ser Glu Pro Asn Pro Leu Xaa Xaa Met Met Asp Asn Ile Arg Lys Arg
                30                  35                  40
gaa act gaa gtg gtc taacactcta taraaaatga acaaaatctc tgaaagcagc       467
Glu Thr Glu Val Val
                45
tcaacctctt ctgaraaaaa aaatatattc tgaggccaac tgttgctaca aaacaaattc      527
tgactgaatg gttaaaacat ttctagtara aggggaaaaa aaakttaaac atgcactgtt      587
tgtgtgtata sccattttcat taaatataca gtaaaactyc aaaaaaaaaa aa .          639
```

```
<210> 277
<211> 772
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 284..463

<221> sig_peptide
<222> 284..379
<223> Von Heijne matrix
      score 3.79999995231628
      seq TFINITLWLGSLC/QR

<221> polyA_site
<222> 762..772

<400> 277
acagctgggg ctttgtcttc tttattgcta ggagaatgta gcaatagaag ttctcatcgc       60
cctgtattgc acttttggtt ttaaggactg gacccagagt tcctgaaagc caaactccat      120
aagctgctca gtaagttcca agcacatagc cggctkhggg atgcgattcg gtcgaggtct      180
gttgaatgaa ggtagacgca gcaggcagtt tgtccttacc agtgacctgg aagacggtgg      240
cacttcctga gtgagctcac ttaccttccc tgaatggtga ggc atg gat gaa tat       295
                                               Met Asp Glu Tyr
                                                        -30
tcc tgg tgg tgc cac gtg tta gag gtg gta aag ggt caa atg ttt act        343
Ser Trp Trp Cys His Val Leu Glu Val Val Lys Gly Gln Met Phe Thr
        -25                  -20                  -15
ttt att aat att aca tta tgg ctt ggt tct ctg tgt cag cga ttt ttc        391
Phe Ile Asn Ile Thr Leu Trp Leu Gly Ser Leu Cys Gln Arg Phe Phe
        -10                  -5                   1
tat gcc tcg ggt act tat ttc cta ata tat atc agc aca gta acg cct        439
Tyr Ala Ser Gly Thr Tyr Phe Leu Ile Tyr Ile Ser Thr Val Thr Pro
5                   10                  15                  20
agc tgg agg ctt tgt ctt gtt agt tgataaatta gtggtaacag gtagatttgg        493
Ser Trp Arg Leu Cys Leu Val Ser
                25
ttacctccca aagtgctggg attrcagacg tgagccaccg cgcctggccg aaacaattct      553
tttgaaagag agaagtctcc ctgtgttgcg caggctggtc tcagactcct ggggtcaagt      613
gagcctcctg ctttcgcctc ctaaagtgct gggattacag gcgtgagcca ccgcacccgg      673
acagatgtgt tgattttaaa gtgggtatga ggcctgagcc ctggagtttg agaccagcct      733
ggacaacatg gcaagaccct gtctctccaa aaaaaaaaa                            772
```

```
<210> 278
<211> 840
<212> DNA
```

<213> Homo sapiens

<220>
<221> CDS
<222> 162..671

<221> sig_peptide
<222> 162..398
<223> Von Heijne matrix
     score 4.09999990463257
     seq QGVLFICFTCARS/FP

<221> polyA_signal
<222> 805..810

<221> polyA_site
<222> 830..840

<400> 278

```
aaaaaactgag gcctgggagc aggaacctgt aggcagcgct tgagggtagc gggatagcag      60
ctgcaacgcg cgtgggaggc gggggctctg ggcggaacaa aaatcacagg atgtcagagg     120
atgtttcccg ggaagaactg ggataaaggg gtcccagcac c atg gag gac ccg aac     176
                                             Met Glu Asp Pro Asn
                                                             -75
cct gaa gag aac atg aag cag cag gat tca ccc aag gag aga agt ccc     224
Pro Glu Glu Asn Met Lys Gln Gln Asp Ser Pro Lys Glu Arg Ser Pro
            -70                 -65                 -60
cag agc cca gga ggc aac atc tgc cac ctg ggg gcc ccg aag tgc acc     272
Gln Ser Pro Gly Gly Asn Ile Cys His Leu Gly Ala Pro Lys Cys Thr
            -55                 -50                 -45
cgc tgc ctc atc acc ttc gca gat tcc aag ttc cag gag cgt cac atg     320
Arg Cys Leu Ile Thr Phe Ala Asp Ser Lys Phe Gln Glu Arg His Met
            -40                 -35                 -30
aag cgg gag cac cca gcg gac ttc gtg gcc cag aag ctg cag ggg gtc     368
Lys Arg Glu His Pro Ala Asp Phe Val Ala Gln Lys Leu Gln Gly Val
    -25                 -20                 -15
ctc ttc atc tgc ttc acc tgc gcc cgc tcc ttc ccc tcc tcc aaa gcc     416
Leu Phe Ile Cys Phe Thr Cys Ala Arg Ser Phe Pro Ser Ser Lys Ala
-10                 -5                  1                   5
ckr rkc acc cac car cgc agc cac ggt cca rcc gcc aag ccc acc ctg     464
Xaa Xaa Thr His Gln Arg Ser His Gly Pro Xaa Ala Lys Pro Thr Leu
            10                  15                  20
ccg gtt gca acc act act gcc car ccc acc ttc cct tgt cct gac tgt     512
Pro Val Ala Thr Thr Thr Ala Gln Pro Thr Phe Pro Cys Pro Asp Cys
    25                  30                  35
ggc aaa acc ttt ggg cag gct gtt tct ctg arg cgg cac csc caa atr     560
Gly Lys Thr Phe Gly Gln Ala Val Ser Leu Xaa Arg His Xaa Gln Xaa
    40                  45                  50
cat gar gtc cgt gcc cct cct ggc acc ttc gcc tgc aca rad tgc ggt     608
His Glu Val Arg Ala Pro Pro Gly Thr Phe Ala Cys Thr Xaa Cys Gly
55                  60                  65                  70
cag gac ttt gct car gaa rca ggg ctg cat caa cac tac att cgg cat     656
Gln Asp Phe Ala Gln Glu Xaa Gly Leu His Gln His Tyr Ile Arg His
            75                  80                  85
gcc cgg ggg gga ctc tgagttcagc ttaagcctct ccacggtgac gggtggctct     711
Ala Arg Gly Gly Leu
            90
gtggctggta ggactcaccc atgatatggg gtgcaggaac tctggggggcc ctgaaggatt     771
tgcttccctc ccctgggaag gcagagggct cttaataaag aggacccaka agattcttaa     831
aaaaaaaaa                                                            840
```

<210> 279
<211> 840
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 63..632

<221> sig_peptide
<222> 63..308
<223> Von Heijne matrix
     score 4.40000009536743
     seq NLPHLQVVGLTWG/HI

<221> polyA_signal
<222> 808..813

<221> polyA_site
<222> 829..840

<400> 279
```
aacttccggt cgcgccascg cccgttgcca gttctgcgcg tgtcctgcat ctccagtatg      60
ga atg tat gtd tgg ccc tgt gct gtg gtc ctg gcc cag tac ctt tgg      107
   Met Tyr Val Trp Pro Cys Ala Val Val Leu Ala Gln Tyr Leu Trp
       -80              -75               -70
ttt cac aga aga tct ctg cca ggc aag gcc atc tta gag att gga gct      155
Phe His Arg Arg Ser Leu Pro Gly Lys Ala Ile Leu Glu Ile Gly Ala
    -65              -60               -55
gga gtg agc ctt cca gga att ttg gct gcc aaa tgt ggt gca gaa gta      203
Gly Val Ser Leu Pro Gly Ile Leu Ala Ala Lys Cys Gly Ala Glu Val
    -50              -45           -40
ata ctg tca gac agc tca gaa ctg cct cac tgt ctg gaa gtc tgt cgg      251
Ile Leu Ser Asp Ser Ser Glu Leu Pro His Cys Leu Glu Val Cys Arg
    -35              -30               -25           -20
caa agc tgc caa atg aat aac ctg cca cat ctg cag gtg gta gga cta      299
Gln Ser Cys Gln Met Asn Asn Leu Pro His Leu Gln Val Val Gly Leu
             -15              -10               -5
aca tgg ggt cat ata tct tgg gat ctt ctg gct cta cca cca caa gat      347
Thr Trp Gly His Ile Ser Trp Asp Leu Leu Ala Leu Pro Pro Gln Asp
             1                5                 10
att atc ctt gca tct gat gtg ttc ttt gaa cca gaa rat ttt gaa gac      395
Ile Ile Leu Ala Ser Asp Val Phe Phe Glu Pro Glu Xaa Phe Glu Asp
    15               20                25
att ttg gct aca ata tat ttt ttg atg cac aar aat ccc aag gtc caa      443
Ile Leu Ala Thr Ile Tyr Phe Leu Met His Lys Asn Pro Lys Val Gln
30               35                40                45
ttg tgg tct act tat caa gtt agg art gct gac tgg tca ctt gaa gct      491
Leu Trp Ser Thr Tyr Gln Val Arg Xaa Ala Asp Trp Ser Leu Glu Ala
             50                55                60
tta ctc tac aaa tgg gat atg aaa tgt gtc cac att cct ctt gag tct      539
Leu Leu Tyr Lys Trp Asp Met Lys Cys Val His Ile Pro Leu Glu Ser
             65                70                75
ttt gat gca gac aaa gaa rat ata gca gaa tct acc ctt cca gga aga      587
Phe Asp Ala Asp Lys Glu Xaa Ile Ala Glu Ser Thr Leu Pro Gly Arg
             80                85                90
cat aca gtt gaa atg ctg gtc att tcc ttt gca aag gac agt ctc      632
His Thr Val Glu Met Leu Val Ile Ser Phe Ala Lys Asp Ser Leu
    95               100               105
tgaattatac ctacaacctg ttctgggaca gtatcaatac tgatgagcaa cctggcacac      692
aaactatgag cagaccactt cagcttgaga atgcagtggg tctgaagatg gtcaagtctg      752
```

```
tttgccttar attttgatgt cacctagaca acacttaaac tcatatgaaa caaaaattaa      812
aatacgtatt acaagcaaaa aaaaaaaa                                         840
```

```
<210> 280
<211> 849
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 21..362

<221> sig_peptide
<222> 21..200
<223> Von Heijne matrix
      score 4.80000019073486
      seq LVILSLKSQTLDA/ET

<221> polyA_signal
<222> 821..826

<221> polyA_site
<222> 838..849

<400> 280
agtaagtccc cccgcctcgc atg atg gct gcg gtg ccg ccg ggc ctg gag ccg      53
                     Met Met Ala Ala Val Pro Pro Gly Leu Glu Pro
                     -60             -55                     -50

tgg aac cgt gtg aga atc cct aag gcg ggg aac cgc agc gca gtg aca      101
Trp Asn Arg Val Arg Ile Pro Lys Ala Gly Asn Arg Ser Ala Val Thr
        -45             -40                     -35

gtg cag aac ccc ggc gcg gcc ctt gac ctt tgc att gca gct gta att      149
Val Gln Asn Pro Gly Ala Ala Leu Asp Leu Cys Ile Ala Ala Val Ile
        -30             -25                 -20

aaa gaa tgc cat ctc gtc ata ctg tcg ctg aag agc caa acc tta gat      197
Lys Glu Cys His Leu Val Ile Leu Ser Leu Lys Ser Gln Thr Leu Asp
    -15             -10                 -5

gca gaa aca gat gtg tta tgt gca gtc ctt tac agc aat cac aac aga      245
Ala Glu Thr Asp Val Leu Cys Ala Val Leu Tyr Ser Asn His Asn Arg
1           5                   10                  15

atg ggc cgc cac aaa ccc cat ttg gcc ctc aaa cag gtt gag caa tgt      293
Met Gly Arg His Lys Pro His Leu Ala Leu Lys Gln Val Glu Gln Cys
            20              25                  30

tta aag cgt ttg aaa aac atg aat ttg gag ggc tca att caa gac ctg      341
Leu Lys Arg Leu Lys Asn Met Asn Leu Glu Gly Ser Ile Gln Asp Leu
        35                  40                  45

ttt gag ttg ttt tct tcc aag taagtaagtg gtccarttgc tttgtgatgt      392
Phe Glu Leu Phe Ser Ser Lys
        50

ggtgggctgg gaactcaatg tcttgtgatc kcccttwgga ttkctctakg ctygckgttg      452
gaatataacc aattataccw cagctgtaka aatwttgttt taatgtgggg taccyggtgt      512
ktgtggtaat cttctgacat tgatctatgg gartgactgg tgtgacattg aaatctgggt      572
catggtagat tatattaaaa catcagtggg ctgttattgt gcttaactac ctcaagttga      632
gcttaaagca agtcttcact tgaaaactgc tatagaaatg ctttatattt aaaaatgaaa      692
gtaatgggar mttgcacata gctgaaaatg tgaagggtcg cccagggagg amatggaagc      752
tctgtgcttc ttctgccata ccttgcccta tgcatctctt tgtttcaatc ctttgtcata      812
tcctttataa taaactggta aatgtaaaaa aaaaaaa                              849
```

```
<210> 281
<211> 1344
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 21..503

<221> sig_peptide
<222> 21..344
<223> Von Heijne matrix
      score 5.30000019073486
      seq ACMTLTASPGVFP/SL

<221> polyA_signal
<222> 1305..1310

<221> polyA_site
<222> 1330..1341

<400> 281
aaacaactcc ggaaagtaca atg acc agc ggg cag gcc cga gct tcc wyc cag          53
                      Met Thr Ser Gly Gln Ala Arg Ala Ser Xaa Gln
                          -105                    -100
tcc ccc cag gcc ctg gag gac tcg ggc ccg gtg aat atc tca gtc tca           101
Ser Pro Gln Ala Leu Glu Asp Ser Gly Pro Val Asn Ile Ser Val Ser
        -95             -90                 -85
atc acc cta acc ctg gac cca ctg aaa ccc ttc gga ggg tat tcc cgc          149
Ile Thr Leu Thr Leu Asp Pro Leu Lys Pro Phe Gly Gly Tyr Ser Arg
    -80             -75                 -70
aac gtc acc cat ctg tac tca acc atc tta ggg cat cag att gga ctt          197
Asn Val Thr His Leu Tyr Ser Thr Ile Leu Gly His Gln Ile Gly Leu
-65             -60                 -55                 -50
tca ggc agg gaa gcc cac gag gag ata aac atc acc ttc acc ctg cct          245
Ser Gly Arg Glu Ala His Glu Glu Ile Asn Ile Thr Phe Thr Leu Pro
            -45                 -40                 -35
aca gcg tgg agc tca gat gac tgc gcc ctc cac ggt cac tgt gag cag          293
Thr Ala Trp Ser Ser Asp Asp Cys Ala Leu His Gly His Cys Glu Gln
        -30                 -25                 -20
gtg gta ttc aca gcc tgc atg acc ctc acg gcc agc cct ggg gtg ttc          341
Val Val Phe Thr Ala Cys Met Thr Leu Thr Ala Ser Pro Gly Val Phe
        -15                 -10                 -5
ccg tca ctg tac agc cac cgc act gtg ttc ctg aca cgt aca gca acg          389
Pro Ser Leu Tyr Ser His Arg Thr Val Phe Leu Thr Arg Thr Ala Thr
    1               5                   10                  15
cca cgc tct ggt aca aga tct tca caa ctg cca gag atg cca aca caa          437
Pro Arg Ser Gly Thr Arg Ser Ser Gln Leu Pro Glu Met Pro Thr Gln
                20                  25                  30
aat acg ccc aaa att aca atc ctt tct ggt gtt ata agg ggg cca ttg          485
Asn Thr Pro Lys Ile Thr Ile Leu Ser Gly Val Ile Arg Gly Pro Leu
                35                  40                  45
gaa aag tct atc atg ctt taaatcccaa gcttacagtg attgttccag                  533
Glu Lys Ser Ile Met Leu
                50
atgatgaccg ttcattaata aatttgcatc tcatgcacac cagttacttc ctctttgtga        593
tggtgataac aatgttttgc tatgctgtta tcaagggcag acctagcaaa ttgcgtcaga        653
gcaatcctga attttgtccc gagaaggtgg ctttggctga agcctaattc cacagctcct        713
tgttttttga gagagactga gagaaccata atccttgcct gctgaaccca gcctgggcct        773
ggatgctctg tgaatacatt atcttgcgat gttgggttat tccagccaaa gacatttcaa        833
gtgcctgtaa ctgatttgta catatttata aaaatctatt cagaaattgg tccaataatg        893
cacgtgcttt gccctgggta cagccagagc ccttcaaccc caccttggac ttgaggacct        953
```

```
acctgatggg acgtttccac gtgtctctag agaaggatcc tggatctagc tggtcacgac    1013
gatgttttca ccaaggtcac aggagcattg cgtcgctgat ggggttgaag tttggtttgg    1073
ttcttgtttc agcccaatat gtagagaaca tttgaaacag tctgcacctt tgatacggta    1133
ttgcattttcc aaagccacca atccattttg tggattttat gtgtctgtgg cttaataatc    1193
atagtaacaa caataatacc tttttctcca ttttgcttgc aggaaacata ccttaagttt    1253
tttttgtttt gtttttgttt ttttgttttt tgtttttcctt tatgaagaaa aaataaaata    1313
gtcacatttt aatacyaaaa aaaaaaaaamc h                                   1344
```

<210> 282
<211> 671
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 1..201

<221> sig_peptide
<222> 1..63
<223> Von Heijne matrix
      score 5.09999990463257
      seq LLLKIWLLQRPES/QE

<221> polyA_signal
<222> 637..642

<221> polyA_site
<222> 660..671

<400> 282
```
atg ctg gga ggt gac cat agg gct ctg ctt tta aag ata tgg ctg ctt     48
Met Leu Gly Gly Asp His Arg Ala Leu Leu Leu Lys Ile Trp Leu Leu
    -20             -15               -10
caa agg cca gag tca cag gaa gga ctt ctt cca ggg aga tta gtg gtg     96
Gln Arg Pro Glu Ser Gln Glu Gly Leu Leu Pro Gly Arg Leu Val Val
    -5              1               5               10
atg gag agg aga gtt aaa aat gac ctc atg tcc ttc ttg tcc acg gtt    144
Met Glu Arg Arg Val Lys Asn Asp Leu Met Ser Phe Leu Ser Thr Val
            15              20              25
ttg ttg agt ttt cac tct tct aat gca agg gtc tca cac tgt gaa cca    192
Leu Leu Ser Phe His Ser Ser Asn Ala Arg Val Ser His Cys Glu Pro
        30              35              40
ctt agg atg tgatcacttt caggtggcca ggaatgttga atgtctttgg           241
Leu Arg Met
        45
ctcagttcat ttaaaaaaga tatctatttg aaagttctca rarttgtaca tatgtttcac   301
agtacaggat ctgtacataa aagtttcttt cctaaaccat tcaccaagag ccaatatcta   361
ggcattttct tggtagcaca aattttctta ttgcttaraa aattgtcctc cttgttattt   421
ctgtttgtaa racttaagtg agttaggtct ttaaggaaag caacgctcct ctgaaatgct   481
tgtctttttt ctgttgccga aatarctggt cctttttcgg gagttaratg tatarartgt   541
ttgtatgtaa acatttcttg taggcatcac catgaacaaa gatatatttt ctatttattt   601
attatatgtg cacttcaaga agtcactgtc agagaaataa agaattgtct taaatgtcaa   661
aaaaaaaaaa                                                           671
```

<210> 283
<211> 1601
<212> DNA
<213> Homo sapiens

```
<220>
<221> CDS
<222> 39..1034

<221> sig_peptide
<222> 39..134
<223> Von Heijne matrix
      score 6.09999990463257
      seq LPLLTSALHGLQQ/QH

<221> polyA_signal
<222> 1566..1571

<221> polyA_site
<222> 1587..1597

<400> 283
agccccagat cctgaaggag gtgcagagcc cagagggg atg atc kcg ctg agg gac          56
                                            Met Ile Xaa Leu Arg Asp
                                                    -30

aca gct gcc tcc ctc cgc ctt gag aga gac aca agg cag ttg cca ctg          104
Thr Ala Ala Ser Leu Arg Leu Glu Arg Asp Thr Arg Gln Leu Pro Leu
    -25             -20                 -15

ctc acc agt gcc ctg cac gga ctg cag cag cag cac cca gcc ttc tct          152
Leu Thr Ser Ala Leu His Gly Leu Gln Gln Gln His Pro Ala Phe Ser
-10                 -5                  1               5

ggt gtg gca cgg ctg gcc aag cgg tgg gtg cgt gcc cag ctt ctt ggt          200
Gly Val Ala Arg Leu Ala Lys Arg Trp Val Arg Ala Gln Leu Leu Gly
            10                  15                  20

gag ggt ttc gct gat gag agc ctg gat ctg gtg gcc gct gcc ctt ttc          248
Glu Gly Phe Ala Asp Glu Ser Leu Asp Leu Val Ala Ala Ala Leu Phe
        25                  30                  35

ctg cac cct gag ccc ttc acc cct ccg agt tcc ccc cag gtt ggc ttc          296
Leu His Pro Glu Pro Phe Thr Pro Pro Ser Ser Pro Gln Val Gly Phe
        40                  45                  50

ctt cga ttc ctt ttc ttg gta tca acg ttt gat tgg aag aac aac ccc          344
Leu Arg Phe Leu Phe Leu Val Ser Thr Phe Asp Trp Lys Asn Asn Pro
55                  60                  65                  70

ctc ttt gtc aac ctc aat aat gag ctc act gtg gag gag cag gtg gar          392
Leu Phe Val Asn Leu Asn Asn Glu Leu Thr Val Glu Glu Gln Val Glu
                75                  80                  85

atc cgc agt ggc ttc ctg gca gct cgg gca cag ctc ccc gtc atg gtc          440
Ile Arg Ser Gly Phe Leu Ala Ala Arg Ala Gln Leu Pro Val Met Val
            90                  95                  100

att gtt acc ccc caa rac cgc aaa aac tct gtg tgg aca cag gat gga          488
Ile Val Thr Pro Gln Xaa Arg Lys Asn Ser Val Trp Thr Gln Asp Gly
            105                 110                 115

ccc tca gcc car atc ctg cag cag ctt gtg gtc ctg gca gct gaa scc          536
Pro Ser Ala Gln Ile Leu Gln Gln Leu Val Val Leu Ala Ala Glu Xaa
    120                 125                 130

ctg ccc atg tta rar aas cag ctc atg gat ccc cgg gga cct ggg gac          584
Leu Pro Met Leu Xaa Xaa Gln Leu Met Asp Pro Arg Gly Pro Gly Asp
135                 140                 145                 150

atc agg aca gkg ttc cgg ccg ccc ttg gac att tac gac gtg ctg att          632
Ile Arg Thr Xaa Phe Arg Pro Pro Leu Asp Ile Tyr Asp Val Leu Ile
                155                 160                 165

cgc ctg tct cct cgc cat atc ccg cgg cac cgc cag gct gtg gac tcr          680
Arg Leu Ser Pro Arg His Ile Pro Arg His Arg Gln Ala Val Asp Ser
            170                 175                 180

cca gct gcc tcc ttc tgc cgg ggc ctg ctc agc cag ccg ggg ccc tca          728
Pro Ala Ala Ser Phe Cys Arg Gly Leu Leu Ser Gln Pro Gly Pro Ser
```

```
              185                  190                  195
tcc ctg atg ccc gtg ctg ggc tak gat cct cct cag ctc tat ctg acg      776
Ser Leu Met Pro Val Leu Gly Xaa Asp Pro Pro Gln Leu Tyr Leu Thr
    200                  205                  210
cag ctc arg gag gcc ttt ggg gat ctg gcc ctt ttc ttc tat gac cag      824
Gln Leu Xaa Glu Ala Phe Gly Asp Leu Ala Leu Phe Phe Tyr Asp Gln
215                  220                  225                  230
cat ggt gga gag gtg att ggt gtc ctc tgg aag ccc acc agc ttc cag      872
His Gly Gly Glu Val Ile Gly Val Leu Trp Lys Pro Thr Ser Phe Gln
                 235                  240                  245
ccg cag ccc ttc aag gcc tcc agc aca aag ggg cgc atg gtg atg tct      920
Pro Gln Pro Phe Lys Ala Ser Ser Thr Lys Gly Arg Met Val Met Ser
             250                  255                  260
cga ggt ggg gag cta gta atg gtg ccc aat gtt gaa gca atc ctg gag      968
Arg Gly Gly Glu Leu Val Met Val Pro Asn Val Glu Ala Ile Leu Glu
         265                  270                  275
gac ttt gct gtg ctg ggt gaa ggc ctg gtg cag act gtg gag gcc cga     1016
Asp Phe Ala Val Leu Gly Glu Gly Leu Val Gln Thr Val Glu Ala Arg
     280                  285                  290
agt gag agg tgg act gtg tgatcccagc tctggagcaa gctgtagacg           1064
Ser Glu Arg Trp Thr Val
295                  300
gacagcagga cattggacct ctagagcaag atgtcagtag gatgacctcc accctccttg   1124
gacatgaatc ctccatggag ggcctgctgg ctgaacatgc tgaatcatct ccaacaaaac   1184
ccagccccaa ctttctctct gatgctccag cattggggca ggggcatggt ggcccatgta   1244
gtctcctggg cctcaccatc ccagaagagg agtgggagcc agctcagaga aggaactgaa   1304
cccaggagat ccatccacct attagccctg ggcctggacc tccctgcgat ttcccactcc   1364
tttcttagtc ttcttccaga aacagagaag gggatgtgtg cctgggagag gctctgtctc   1424
cttcctgctg ccaggacctg tgcctagact tagcatgccc ttcactgcag tgtcaggcct   1484
ttagatggga cccagcgaaa atgtggccct tctgagtcac atcaccgaca ctgagcagtg   1544
gaaagggggct atatgtgtat gaatagacca cattgaagga gcaaaaaaaa aaamcch      1601
```

```
<210> 284
<211> 1206
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 69..263

<221> sig_peptide
<222> 69..125
<223> Von Heijne matrix
      score 3.90000009536743
      seq ALSMSSFSFHSSS/CS

<221> polyA_signal
<222> 1173..1178

<221> polyA_site
<222> 1196..1205

<400> 284
acatttgtga ctttaccaat accctcccag ttcttgatag acagctgtag gttgctgggt     60
tcaagaat atg ggt ggg ata tgg aat gct ctt tca atg tct agc ttc agt    110
         Met Gly Gly Ile Trp Asn Ala Leu Ser Met Ser Ser Phe Ser
             -15                  -10
ttt cat tca tcc tcc tgc tca gca ctg tca gcc aag agc tta ctc agc    158
Phe His Ser Ser Ser Cys Ser Ala Leu Ser Ala Lys Ser Leu Leu Ser
```

```
     -5                   1                   5                   10
aga cac cac ata ctg cag cag ttc cta gtg aga aaa tct gtg cca cta      206
Arg His His Ile Leu Gln Gln Phe Leu Val Arg Lys Ser Val Pro Leu
                15                  20                  25
gaa aat gct tca ctt cca ttt cct cac ctg ggc agt tct ctg ttt aaa      254
Glu Asn Ala Ser Leu Pro Phe Pro His Leu Gly Ser Ser Leu Phe Lys
        30                  35                  40
att gtg ggc tgatttggtc ttcctctcct cctcccactg ttactgccct             303
Ile Val Gly
        45

gcagcccttg ttcaggtgta cagaccctta ttctggcctc tagtgtcctt gtctgtcatg   363
acacaccctt ccgcccaaat acctctgacc ccaaggctgg aatggggctg gtaggarata   423
agtttgctta ctcatartca tgtcctttct cttggcacct gcttccctgc ggtgtcctca   483
aatggatttc tgtgtggcag tggartgatt gcatgaattt ttctgtaaca cattaacttt   543
gtattattat taagggartt tgaraaagct ttgcttataa tgtcaaggca aggaggtaaa   603
aactggagcc caaakaaatt cccttagggc aagattatgt tataataraa aattgaattt   663
cctgaggcag tggctgccac cccttttcar atgtttagtc ctgcaaatag catctttctt   723
gtag_ctgtg acatggatgg ggatgctagg gcccttaggg gcaaggggac taaactaaat   783
caakttgagt tttttttccag caggggttar gggaggtact cctgttgat atttgacact    843
araaagtaat ctttttttaca aaactgtttt tctaggtggg tggaaagtga aactgccaca   903
tccttgttgg tttagtccaa raratcattt gcaacaacag taratgtccg ggttttgttt   963
ctgtcttttt attatgaaaa actatgttaa gggggaaaat gtggattatg gtaaccarag   1023
gaatccctas ccttgttttc cttaraaaac ttgtttagtg ttttatcara cgtctgttgt   1083
agttgtarac aggaaagctt gtgaraaaaa caccacatgg ascctgtaaa tgttttttgca   1143
caacctgtaa agcattcttg gaaktggcca gtaaaaaggg gttttaccat ttaaaaaaaa   1203
aat                                                                  1206
```

```
<210> 285
<211> 536
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 115..285

<221> sig_peptide
<222> 115..204
<223> Von Heijne matrix
      score 3.70000004768372
      seq SMMLLTVYGGYLC/SV

<221> polyA_signal
<222> 505..510

<221> polyA_site
<222> 525..536

<400> 285
acgagtgctg cgttcggctg tgctgggaag ttgcgtagac agtggcctcg agaccctgcc   60
tgcctgagga ggcctcggtt ggatgcgaag gagctgcagc atccagggga caag atg     117
                                                                Met
                                                                -30
cca act ggc aag cag cta gct gac att ggc tat aag acc ttc tct acc     165
Pro Thr Gly Lys Gln Leu Ala Asp Ile Gly Tyr Lys Thr Phe Ser Thr
            -25                 -20                 -15
tcc atg atg ctt ctc act gtg tat ggg ggg tac ctc tgc agt gtc cga     213
Ser Met Met Leu Leu Thr Val Tyr Gly Gly Tyr Leu Cys Ser Val Arg
            -10                 -5                  1
gtc tac cac tat ttc cag tgg cgc agg gcc cag cgc cag gcc gca gaa     261
```

```
Val Tyr His Tyr Phe Gln Trp Arg Arg Ala Gln Arg Gln Ala Ala Glu
    5                   10                  15
gaa cag aag dac tca gga atc atg tagaactggg gggctttttc tcctgagcar     315
Glu Gln Lys Xaa Ser Gly Ile Met
20                  25
asakgcccaa ggcatgctgt ggagagactt cacctgccac catttccagg tcaacaggac     375
tagagcgttg atggttttca aaccctgttg gaagaaagtg cccatggttt ctctggttct     435
gccartttga cagtttatgg argcttttga atcgtaatar caatgtgagg gtgargtaca     495
cctacagaca ttaaataatt tgctgtgtca aaaaaaaaa a·                          536
```

```
<210> 286
<211> 529
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 90..344

<221> sig_peptide
<222> 90..140
<223> Von Heijne matrix
      score 8.19999980926514
      seq LLLITAILAVAVG/FP

<221> polyA_signal  ·
<222> 500..505

<221> polyA_site
<222> 515..527
```

```
<400> 286
aatatrarac agctacaata ttccagggcc artcacttgc catttctcat aacagcgtca      60
gagagaaaga actgactgar acgtttgag atg aag aaa gtt ctc ctc ctg atc       113
                                  Met Lys Lys Val Leu Leu Leu Ile
                                  -15                         -10
aca gcc atc ttg gca gtg gct gtw ggt ttc cca gtc tct caa gac cag       161
Thr Ala Ile Leu Ala Val Ala Val Gly Phe Pro Val Ser Gln Asp Gln
            -5                  1               5
gaa cga gaa aaa aga agt atc agt gac agc gat gaa tta gct tca ggr       209
Glu Arg Glu Lys Arg Ser Ile Ser Asp Ser Asp Glu Leu Ala Ser Gly
        10                  15                  20
wtt ttt gtg ttc cct tac cca tat cca ttt cgc cca ctt cca cca att       257
Xaa Phe Val Phe Pro Tyr Pro Tyr Pro Phe Arg Pro Leu Pro Pro Ile
    25                  30                  35
cca ttt cca aga ttt cca tgg ttt aga cgt aat ttt cct att cca ata       305
Pro Phe Pro Arg Phe Pro Trp Phe Arg Arg Asn Phe Pro Ile Pro Ile
40                  45                  50                  55
cct gaa tct gcc cct aca act ccc ctt cct agc gaa aag taaacaaraa        354
Pro Glu Ser Ala Pro Thr Thr Pro Leu Pro Ser Glu Lys
            60                  65
ggaaaagtca crataaacct ggtcacctga aattgaaatt gagccacttc cttgaaraat     414
caaaattcct gttaataaaa raaaaacaaa tgtaattgaa atagcacaca gcattctcta     474
gtcaatatct ttagtgatct tctttaataa acatgaaagc aaaaaaaaaa aaacc          529
```

```
<210> 287
<211> 493
<212> DNA
```

```
<213> Homo sapiens

<220>
<221> CDS
<222> 57..311

<221> sig_peptide
<222> 57..107
<223> Von Heijne matrix
      score 8.19999980926514
      seq LLLITAILAVAVG/FP

<221> polyA_signal
<222> 467..472

<221> polyA_site
<222> 482..493

<400> 287
aacttgccat ttctcataac agcgtcagag agaaagaact gactgaaacg tttgag atg     59
                                                                 Met

aag aaa gtt ctc ctc ctg atc aca gcc atc ttg gca gtg gct gtt ggt     107
Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val Gly
    -15                 -10                 -5

ttc cca gtc tct caa gac cak gaa cga gaa aaa aga agt atc agt gac     155
Phe Pro Val Ser Gln Asp Xaa Glu Arg Glu Lys Arg Ser Ile Ser Asp
1                   5                   10                  15

agc gat gaa tta gct tca ggg ttt ttt gtg ttc cct tac cca tat cca     203
Ser Asp Glu Leu Ala Ser Gly Phe Phe Val Phe Pro Tyr Pro Tyr Pro
            20                  25                  30

ttt cgc cca ctt cca cca att cca ttt cca aga ttt cca tgg ttt aga     251
Phe Arg Pro Leu Pro Pro Ile Pro Phe Pro Arg Phe Pro Trp Phe Arg
        35                  40                  45

cgt aat ttt cct att cca ata cct gaa tct gcc cct aca act ccc ctt     299
Arg Asn Phe Pro Ile Pro Ile Pro Glu Ser Ala Pro Thr Thr Pro Leu
    50                  55                  60

ccg agc gaa aag taaacaagaa ggaaaagtca cgataaacct ggtcacctga        351
Pro Ser Glu Lys
65
aattgaaatt gagccacttc cttgargaat caaaattcct gttaataaaa gaaaaacaaa    411
tgtaattgaa atagcacaca gcattctcta gtcaatatct ttagtgatct tctttaataa    471
acatgaaagc aaaaaaaaaa aa                                             493
```

```
<210> 288
<211> 521
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 96..302

<221> sig_peptide
<222> 96..182
<223> Von Heijne matrix
      score 5
      seq ELSLLPSSLWVLA/TS

<221> polyA_site
<222> 501..514
```

```
<400> 288
aagagacgtc accggctgcg cccttcagta tcgcggacgg aagatggcgt ccgccacccg      60
tctcatccag cggctgcgga actgggcgtc cgggc atg acc tgc agg gga agc       113
                                     Met Thr Cys Arg Gly Ser
                                                          -25
tgc agc tac gct acc agg aga tct cca agc gaa ctc agc ctc ctc cca     161
Cys Ser Tyr Ala Thr Arg Arg Ser Pro Ser Glu Leu Ser Leu Leu Pro
            -20                 -15                 -10
agc tcc ctg tgg gtc cta gcc aca agc tct cca aca att act att gca     209
Ser Ser Leu Trp Val Leu Ala Thr Ser Ser Pro Thr Ile Thr Ile Ala
        -5                   1               5
ctc gcg atg gcc gcc ggg aat ctg tgc ccc ctt cca tca tca tkt cgt     257
Leu Ala Met Ala Ala Gly Asn Leu Cys Pro Leu Pro Ser Ser Xaa Arg
10                  15                  20                  25
crc aaa agg cgc tgg tgt cag gca asc car caa ara gct ctg ctg         302
Xaa Lys Arg Arg Trp Cys Gln Ala Xaa Gln Gln Xaa Ala Leu Leu
                30                  35                  40
tagctgccac tgaaaaraag gcggtgactc cagctcctcc cataaagagg tgggagctgt     362
cctcggacca gccttacctg tgacactgca ccctcacggc cacccgacta ctttgcctcc     422
ttggatttcc tccagggaga atgtgaccta atttatgaca aatacgtara gctcaggtat     482
cacttctagt tttactttaa aaaataaaaa aatagagac                           521
```

```
<210> 289
<211> 811
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 161..526

<221> sig_peptide
<222> 161..328
<223> Von Heijne matrix
      score 4.19999980926514
      seq XSPLLTLALLGQC/SL

<221> polyA_site
<222> 799..811

<400> 289
aaaaaattgc agtgctgaag acactggacc cgcaaaaggc tgtccctccc aaacctggga     60
ttctgggctc actgagttca cctgcgagtc agccctacct gcactgctct ggtctagtac     120
aaacaggctg ctggcattga ggtctgctac aaaaanarta atg gtc cca tgg ccc      175
                                            Met Val Pro Trp Pro
                                                              -55
agg ggc aag gtg aaa act gct cct att ccc atc tct agg ttt cct ttc     223
Arg Gly Lys Val Lys Thr Ala Pro Ile Pro Ile Ser Arg Phe Pro Phe
    -50                 -45                 -40
ctc cct acc cac gac cca ccc acc cca gca cat tgg tct cca gca tct     271
Leu Pro Thr His Asp Pro Pro Thr Pro Ala His Trp Ser Pro Ala Ser
-35                 -30                 -25                 -20
cat cag cag ttt aaa cat kkg tca ccc ctc ctc act ttg gcc ctg ctg     319
His Gln Gln Phe Lys His Xaa Ser Pro Leu Leu Thr Leu Ala Leu Leu
            -15                 -10                 -5
ggt cag tgc tct ctg ttc arc aat ttg agg aaa aaa ctt gca ggg caa     367
Gly Gln Cys Ser Leu Phe Xaa Asn Leu Arg Lys Lys Leu Ala Gly Gln
            1               5                   10
aaa gca aaa aaa tta cct tcc ttc tcc agc ctg ccc ctg aca ctc tgg     415
```

```
Lys Ala Lys Lys Leu Pro Ser Phe Ser Ser Leu Pro Leu Thr Leu Trp
    15                  20                  25
cca tta act cct caa ttt gct gag ctc act aca gtg gca caa aaa aaa    463
Pro Leu Thr Pro Gln Phe Ala Glu Leu Thr Thr Val Ala Gln Lys Lys
30                  35                  40                  45
ttg agg tgg tcc ggg acc cta ggt tgg ggt cca gtt ccc agc tgg gtt    511
Leu Arg Trp Ser Gly Thr Leu Gly Trp Gly Pro Val Pro Ser Trp Val
                50                  55                  60
caa ttt ttt tta ggg tgaatggagg garagttggg gactgaaaas ccttcaaara    566
Gln Phe Phe Leu Gly
                65
caatgttatt acagcaktct ccccttatcc aaaktttcct tttcctgadt ttcagttagc    626
tatggtcaac cgcttggaaa atakttgaac acagtacaat aaratatttt gaggctggga    686
ktggtggctc atgcctgtaa taatcccagg actttgtgar accaaktttg aaggatcact    746
tgaacccagg aktttgarac cascctgggc aacatrgtra gacctcatct ctacaaaaaa    806
aaaaa                                                               811
```

```
<210> 290
<211> 625
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 210..332

<221> sig_peptide
<222> 210..299
<223> Von Heijne matrix
      score 8.10000038146973
      seq ITCLLAFWVPASC/IQ

<221> polyA_signal
<222> 594..599

<221> polyA_site
<222> 613..625

<400> 290
acaggtcsmc ttaacatctc ttgatttgag ccactcccac tgtcatcagc tttcacctgg    60
attatcgtga cagcctccta ctgcttctct atcatgtggc cagagctatc ttccctaaaa   120
atgcattgca tagttgatca agtcactctc tggcctaaaa ccttccttgg ctccctgctg   180
ccctcaggat aaagtctgga cccctcagc atg gct tgt gag act cat ggt gtc    233
                                   Met Ala Cys Glu Thr His Gly Val
                                   -30                 -25
ctt gtc cct gct cac ctc tct ggt ctc atc act tgc ctt ctt gca ttc    281
Leu Val Pro Ala His Leu Ser Gly Leu Ile Thr Cys Leu Leu Ala Phe
        -20                 -15                 -10
tgg gtc cca gcc tcc tgt atc cag aga tgc agt ggc tct cca ttg cca    329
Trp Val Pro Ala Ser Cys Ile Gln Arg Cys Ser Gly Ser Pro Leu Pro
    -5                  1                   5                   10
ctc tgattcctcc tttcttttgg tcacagagaa agggtacttt ctctgtcaaa          382
Leu
tctcaactta gacttgactt cctccaagga gctttggcta tactctctcc cwcgacccc    442
accctggcat actacacara tcactctggg ctcacttgcc tgcctaatgg tcatctcccc    502
agtaaactgt aagctccttg agggcaagga ttgtgttgga atttttgtat taacagtgcc    562
tggcttggtg cctggcacct aaaaagcact caataaatgt ttgtttaatg aaaaaaaaaa    622
aaa                                                                 625
```

<210> 291
<211> 684
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 212..361

<221> sig_peptide
<222> 212..319
<223> Von Heijne matrix
      score 4.09999990463257
      seq HWLFLASLSGIKT/YQ

<221> polyA_signal
<222> 650..655

<221> polyA_site
<222> 673..684

<400> 291
atccccawns cactctctca cagagactgt tcttttcctt ctgagaccct actccagctt   60
gtagttctaa atctgtgatt atgcactgtc tgtcttcctc ttgaggtcag gggccatttc   120
ttttgttctc tgctatgctc aggacccaga tcaaaggagc tcagtaacta tttacaggcg   180
tacatcatat gtggaggaca cttatgctgt g atg gcc cca cac aca gct tcc   232
                                   Met Ala Pro His Thr Ala Ser
                                   -35   -30
ttt ggg gtc tgt ccc ctg ctc tcc gtt acc cgc gtg gta gcc act gag   280
Phe Gly Val Cys Pro Leu Leu Ser Val Thr Arg Val Val Ala Thr Glu
            -25   -20   -15
cac tgg ctc ttc ctg gct tca ctc tct ggc atc aaa act tat cag tcc   328
His Trp Leu Phe Leu Ala Ser Leu Ser Gly Ile Lys Thr Tyr Gln Ser
        -10   -5   1
tac atc tca gtc ttt tgc aag gtg aca ctt atc tgattaccta attcacacra   381
Tyr Ile Ser Val Phe Cys Lys Val Thr Leu Ile
        5   10
aggtgttaat ggtggtaatg gcataktatt tattacccca ggggacccak aacggtggta   441
tcaaaacata tcattcccca gtggtttaaa actctggtag ctttccargg aatccaaagt   501
ggaatccagt ctccttagct gawttcacag ggccccgtct gcacaacttg gcttctgtcg   561
gcttccctan ccctgacttc ccaagcctta gtcatcaccc tctctcccac ccagggctca   621
gcacagtacc tggaacagtc aagccctcaa taaatgttta ctgagtgcat yaaaaaaaaa   681
aaa   684

<210> 292
<211> 628
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 75..482

<221> sig_peptide
<222> 75..128
<223> Von Heijne matrix
      score 3.59999990463257
      seq KMLISVAMLGAXA/GV

<221> polyA_signal
<222> 595..600

<221> polyA_site
<222> 618..627

<400> 292
aagtgagacc gcgcggcaac agcttgcggc tgcggggagc tcccgtgggc gctccgctgg   60
ctgtgcaggc ggcc atg gat tcc ttg cgg aaa atg ctg atc tca gtc gca   110
              Met Asp Ser Leu Arg Lys Met Leu Ile Ser Val Ala
                      -15   -10
atg ctg ggc gca rgg gct ggc gtg ggc tac gcg ctc ctc gtt atc gtg   158
Met Leu Gly Ala Xaa Ala Gly Val Gly Tyr Ala Leu Leu Val Ile Val
    -5   1   5   10
acc ccg gga gag cgg cgg aag cag gaa atg cta aag gag atg cca ctg   206
Thr Pro Gly Glu Arg Arg Lys Gln Glu Met Leu Lys Glu Met Pro Leu
        15   20   25
cag gac cca agg agc agg gag gag gcg gcc agg acc cag cag ct: ttg   254
Gln Asp Pro Arg Ser Arg Glu Glu Ala Ala Arg Thr Gln Gln Leu Leu
        30   35   40
ctg gcc act ctg cag gag gca gcg acc acg cag gag aac gtg gcc tgg   302
Leu Ala Thr Leu Gln Glu Ala Ala Thr Thr Gln Glu Asn Val Ala Trp
        45   50   55
agg aag aac tgg atg gtt ggc ggc gaa ggc ggc gcc acg gga kgt cac   350
Arg Lys Asn Trp Met Val Gly Gly Glu Gly Gly Ala Thr Gly Xaa His
    60   65   70
cgt gag acc gga ctt gcc tcc gtg ggc gcc gga cct tgg ctt ggg cgc   398
Arg Glu Thr Gly Leu Ala Ser Val Gly Ala Gly Pro Trp Leu Gly Arg
75   80   85   90
agg aat ccg agg cag ctt tct cct tcg tgg gcc can cgg aaa atc cgg   446
Arg Asn Pro Arg Gln Leu Ser Pro Ser Trp Ala Xaa Arg Lys Ile Arg
        95   100   105
amc gaa aat wcc atg cca gga ctc tcc ggg gtc ctg tgaactgccg   492
Xaa Glu Asn Xaa Met Pro Gly Leu Ser Gly Val Leu
        110   115
tcgggtgagc acgtgtcccc caaaccctgg actgactgct ttaaggtccg caaggcgggc   552
cagggccgag acgcgagtcg datgtggtga actgaaagaa ccaataaaat catgttcctc   .612
cammcaaaaa aaaaah   628

<210> 293
<211> 813
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 50..631

<221> sig_peptide
<222> 50..244
<223> Von Heijne matrix
      score 8
      seq LTLIGCLVTGVES/KI

<221> polyA_signal
<222> 777..782

<221> polyA_site
<222> 801..812

<400> 293
```
aaggaaagga ttactcgagc cttgttagaa tcagacatgg cttcagggg atg cag gac      58
                                                           Met Gln Asp
                                                           -65

gct ccc ctg agc tgc ctg tca ccg act aag tgg agc agt gtt tct tcc      106
Ala Pro Leu Ser Cys Leu Ser Pro Thr Lys Trp Ser Ser Val Ser Ser
        -60             -55                 -50

gca gac tca act gag aag tca gcc tct gcg gca ggc acc agg aat ctg      154
Ala Asp Ser Thr Glu Lys Ser Ala Ser Ala Ala Gly Thr Arg Asn Leu
    -45                 -40                 -35

cct ttt cag ttc tgt ctc cgg cag gct ttg agg atg aag gct gcg ggc      202
Pro Phe Gln Phe Cys Leu Arg Gln Ala Leu Arg Met Lys Ala Ala Gly
-30                 -25                 -20                 -15

att ctg acc ctc att ggc tgc ctg gtc aca ggc gtc gag tcc aaa atc      250
Ile Leu Thr Leu Ile Gly Cys Leu Val Thr Gly Val Glu Ser Lys Ile
                -10             -5                  1

tac act cgt tgc aaa ctg gca aaa ata ttc tcg agg gct ggc ctg gac      298
Tyr Thr Arg Cys Lys Leu Ala Lys Ile Phe Ser Arg Ala Gly Leu Asp
        5               10                  15

aat cyg agg ggc ttc agc ctt gga aac tgg atc tgc atg gcg tat tat      346
Asn Xaa Arg Gly Phe Ser Leu Gly Asn Trp Ile Cys Met Ala Tyr Tyr
    20              25                  30

gag agc ggc tac aac acc aca gcc car acg gtc ctg gat gac ggc agc      394
Glu Ser Gly Tyr Asn Thr Thr Ala Gln Thr Val Leu Asp Asp Gly Ser
35                  40                  45                  50

atc gac tay ggc atc ttc caa atc aac agc ttc gcg tgg tgc aga cgc      442
Ile Asp Tyr Gly Ile Phe Gln Ile Asn Ser Phe Ala Trp Cys Arg Arg
                55                  60                  65

gga aag ctg aag gag aac aac cac tgc cay gtc gcc tgc tca gcc ttg      490
Gly Lys Leu Lys Glu Asn Asn His Cys His Val Ala Cys Ser Ala Leu
            70                  75                  80

rtc act gat gac ctc aca gat gca att atc tgt gcc arg aaa att gtt      538
Xaa Thr Asp Asp Leu Thr Asp Ala Ile Ile Cys Ala Xaa Lys Ile Val
        85                  90                  95

aaa gag aca caa gga atg aac tat tgg caa ggc tgg aag aaa cay tgt      586
Lys Glu Thr Gln Gly Met Asn Tyr Trp Gln Gly Trp Lys Lys His Cys
        100                 105                 110

gag ggg aga gac ctg tcc gas tgg aaa aaa ggc tgt gag gtt tcc          631
Glu Gly Arg Asp Leu Ser Xaa Trp Lys Lys Gly Cys Glu Val Ser
115                 120                 125

taaactggaa ctggacccag gatgctttgc ascaacgccc tagggtttgc agtgaatgtc    691
caaatgcctg tgtcatcttg tcccgtttcc tcccaatatt ccttctcaaa cttggagagg    751
gaaaattaag ctatactttt aagaaaataa atatttccat ttaaatgtca amaaaaaaaa    811
ah                                                                    813
```

<210> 294
<211> 778
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 154..576

<221> sig_peptide
<222> 154..360
<223> Von Heijne matrix
      score 4.80000019073486
      seq MMVLSLGIILASA/SF

```
<221> polyA_signal
<222> 737..742

<221> polyA_site
<222> 763..775

<400> 294
agtaaaaaaa cactggaata aggaagggct gatgactttc agaagatgaa ggtaagtaga    60
aaccgttgat gggactgaga aaccagagtk aaaacctctt tggagcttct gaggactcag   120
ctggaaccaa cgggcacagt tggcaacacc atc atg aca tca caa cct gtt ccc   174
                                   Met Thr Ser Gln Pro Val Pro
                                                        -65
aat gag acc atc ata gtg ctc cca tca aat gtc atc aac ttc tcc caa    222
Asn Glu Thr Ile Ile Val Leu Pro Ser Asn Val Ile Asn Phe Ser Gln
         -60                 -55                 -50
gca gag aaa ccc gaa ccc acc aac cag ggg cag gat agc ctg aag aaa    270
Ala Glu Lys Pro Glu Pro Thr Asn Gln Gly Gln Asp Ser Leu Lys Lys
     -45                 -40                 -35
cat cta cac gca gaa atc aaa gtt att ggg act atc cag atc ttg tgt    318
His Leu His Ala Glu Ile Lys Val Ile Gly Thr Ile Gln Ile Leu Cys
-30                 -25                 -20                 -15
ggc atg atg gta ttg agc ttg ggg atc att ttg gca tct gct tcc ttc    366
Gly Met Met Val Leu Ser Leu Gly Ile Ile Leu Ala Ser Ala Ser Phe
             -10                 -5                   1
tct cca aat ttt acc caa gtg act tct aca ctg ttg aac tct gct tac    414
Ser Pro Asn Phe Thr Gln Val Thr Ser Thr Leu Leu Asn Ser Ala Tyr
         5                   10                  15
cca ttc ata gga ccc ttt ttt gtr akt aaa btt tct gag gag ggc agg    462
Pro Phe Ile Gly Pro Phe Phe Val Xaa Lys Xaa Ser Glu Glu Gly Arg
     20                  25                  30
atg ggg caa ara ggg gag gaa rat vcc aat agc tta aac ttc cca sct    510
Met Gly Gln Xaa Gly Glu Glu Xaa Xaa Asn Ser Leu Asn Phe Pro Xaa
35                  40                  45                  50
gcc agc ttg cta tkt ttg atc tgc cag gav caa gga ttc aac ggt gaa    558
Ala Ser Leu Leu Xaa Leu Ile Cys Gln Xaa Gln Gly Phe Asn Gly Glu
                 55                  60                  65
tct tgt tct cct gtc ggg targataaca ggggttgctt rattttagat          606
Ser Cys Ser Pro Val Gly
                 70
caatttctta tcagactcaa ataaacattt cttttgaaaa tcatcttatt cttcacatta   666
tcatcttgag ctatgatgga aactagtgas ktctctccag gtttaggcga aaaaaaaatc   726
catgaattag gataaagttg ggaaggaaca ttttatacaa aaaaaaaah cc            778
```

```
<210> 295
<211> 1060
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 154..897

<221> sig_peptide
<222> 154..360
<223> Von Heijne matrix
      score 4.80000019073486
      seq MMVLSLGIILASA/SF

<221> polyA_signal
<222> 1017..1022
```

<221> polyA_site
<222> 1044..1054

<400> 295
agtaaaaaaa cactggaata aggaagggct gatgactttc agaagatgaa ggtaagtaga   60
aaccgttgat gggactgaga aaccagagtk aaaacctctt tggagcttct gaggactcag   120
ctggaaccaa cgggcacagt tggcaacacc atc atg aca tca caa cct gtt ccc   174
                                    Met Thr Ser Gln Pro Val Pro
                                                            -65
aat gag acc atc ata gtg ctc cca tca aat gtc atc aac ttc tcc caa   222
Asn Glu Thr Ile Ile Val Leu Pro Ser Asn Val Ile Asn Phe Ser Gln
        -60   -55   -50
gca gag aaa ccc gaa ccc acc aac cag ggg cag gat agc ctg aag aaa   270
Ala Glu Lys Pro Glu Pro Thr Asn Gln Gly Gln Asp Ser Leu Lys Lys
    -45   -40   -35
cat cta cac gca gar rtc aaa gtt att ggg act atc cag atc ttg tgt   318
His Leu His Ala Glu Xaa Lys Val Ile Gly Thr Ile Gln Ile Leu Cys
-30   -25   -20   -15
ggc atg atg gta ttg agc ttg ggg atc att ttg gca tct gct tcc ttc   366
Gly Met Met Val Leu Ser Leu Gly Ile Ile Leu Ala Ser Ala Ser Phe
                -10   -5   1
tct cca aat ttt acc caa gtg act tct aca ctg ttg aac tct gct tac   414
Ser Pro Asn Phe Thr Gln Val Thr Ser Thr Leu Leu Asn Ser Ala Tyr
        5   10   15
cca ttc ata gga ccc ttt ttt ttt atc atc tct ggc tct cta tca atc   462
Pro Phe Ile Gly Pro Phe Phe Phe Ile Ile Ser Gly Ser Leu Ser Ile
    20   25   30
gcc aca aaa aaa agg tta acc aac ctt ttg gtg cat acc acc ctg gtt   510
Ala Thr Lys Lys Arg Leu Thr Asn Leu Leu Val His Thr Thr Leu Val
35   40   45   50
gga agc att ctg agt gct ctg tct gcc ctg gtg ggt ttc att ayc ctg   558
Gly Ser Ile Leu Ser Ala Leu Ser Ala Leu Val Gly Phe Ile Xaa Leu
                55   60   65
tct gtc aaa cag gcc acc tta aat cct gcc tca ctg cak tgt gag ttg   606
Ser Val Lys Gln Ala Thr Leu Asn Pro Ala Ser Leu Xaa Cys Glu Leu
        70   75   80
gmc aaa aat aat ata cca aca ara akt tat gtt yct tac ttt tat cat   654
Xaa Lys Asn Asn Ile Pro Thr Xaa Xaa Tyr Val Xaa Tyr Phe Tyr His
        85   90   95
gat tca ctt tat acc acg gac kgc tat aca gcc aaa gcc akt ctg gct   702
Asp Ser Leu Tyr Thr Thr Asp Xaa Tyr Thr Ala Lys Ala Xaa Leu Ala
    100   105   110
gga act ctc tct ctg atg ctg att tgc act ctg ctg gaa ttc tgc cwa   750
Gly Thr Leu Ser Leu Met Leu Ile Cys Thr Leu Leu Glu Phe Cys Xaa
115   120   125   130
sct gtg ctc act gct gtg ctg cgg tgg aaa cag gct tac tct gac ttc   798
Xaa Val Leu Thr Ala Val Leu Arg Trp Lys Gln Ala Tyr Ser Asp Phe
                135   140   145
cct ggg agt gta ctt ttc ctg cct cam agt tac att ggw aat tct ggm   846
Pro Gly Ser Val Leu Phe Leu Pro Xaa Ser Tyr Ile Gly Asn Ser Gly
                150   155   160
atg tcc tca aaa atg acy cat gac tgt gga tat gaa gaa cta ttg act   894
Met Ser Ser Lys Met Thr His Asp Cys Gly Tyr Glu Glu Leu Leu Thr
                165   170   175
tct taagaaaaaa gggagaaata ttaatcagaa agttgattct tatgataata   947
Ser
tggaaaagtt aaccattata gaaaagcaaa gcttgagttt cctaaatgta agcttttaaa   1007
gtaatgaaca ttaaaaaaaa ccattatttc actgtcaaaa aaaaaaamcc nkt   1060

```
<210> 296
<211> 444
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 146..292

<221> sig_peptide
<222> 146..253
<223> Von Heijne matrix
      score 5.5
      seq FTSMCILFHCLLS/FQ

<221> polyA_signal
<222> 395..400

<221> polyA_site
<222> 433..444

<400> 296
aacttgggac aagaratcaa actttaaaga tggtctaaag cccctcttaa aggtctgact      60
gtgtcggacc tctagagcta atctcactag atgtgagcca ttgtttatat tctagccatc     120
ctttcatttc attctagaag acccc atg caa gtt ccc cac cta agg gtc tgg       172
                              Met Gln Val Pro His Leu Arg Val Trp
                              -35                    -30
aca cag gtg awa gat acc ttc att ggt tat aga aat ttg gga ttt aca       220
Thr Gln Val Xaa Asp Thr Phe Ile Gly Tyr Arg Asn Leu Gly Phe Thr
    -25                 -20                     -15
agt atg tgc ata ttg ttc cac tgt ctt ctt agc ttt cag gtt ttc aaa       268
Ser Met Cys Ile Leu Phe His Cys Leu Leu Ser Phe Gln Val Phe Lys
    -10             -5                  1                   5
aag aaa aga aaa ctt ara ctt ttc tgatgttctt ttttacgtaa ataaccattt      322
Lys Lys Arg Lys Leu Xaa Leu Phe
                10
tattgttgtt ttgctttttc tgccttcaaa ctactcccac aggccaaata tavctggctg     382
cttctttctg taaataaagt tttattgggc cacagccatg gccatctttt aaaaaaaaaa     442
aa                                                                    444
```

```
<210> 297
<211> 754
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 126..383

<221> sig_peptide
<222> 126..167
<223> Von Heijne matrix
      score 7.5
      seq VALNLILVPCCAA/WC

<221> polyA_signal
<222> 726..731

<221> polyA_site
<222> 743..754
```

```
<400> 297
aattgtatgt tacgatgttg tattgatttt taagaaagta attkratttg taaaacttct          60
gctcgtttac actgcacatt gaatacaggt aactaattgg wwggagaggg gaggtcactc         120
ttttg atg gtg gcc ctg aac ctc att ctg gtt ccc tgc tgc gct gct tgg        170
      Met Val Ala Leu Asn Leu Ile Leu Val Pro Cys Cys Ala Ala Trp
             -10                        -5                          1
tgt gac cca cgg agg atc cac tcc cag gat gac gtg ctc cgt agc tct          218
Cys Asp Pro Arg Arg Ile His Ser Gln Asp Asp Val Leu Arg Ser Ser
             5                         10                  15
gct gct gat act ggg tct gcg atg cag cgg cgt gag gcc tgg gct ggt          266
Ala Ala Asp Thr Gly Ser Ala Met Gln Arg Arg Glu Ala Trp Ala Gly
             20                        25                  30
tgg aga agg tca caa ccc ttc tct gtt ggt ctg cct tct gct gaa aga         314
Trp Arg Arg Ser Gln Pro Phe Ser Val Gly Leu Pro Ser Ala Glu Arg
       35                        40                  45
ctc gag aac caa cca ggg aag ctg tcc tgg agg tcc ctg gtc gga gag         362
Leu Glu Asn Gln Pro Gly Lys Leu Ser Trp Arg Ser Leu Val Gly Glu
50                        55                  60                  65
gga cat aga atc tgt gac ctc tgacrrctgt gaasccaccc tgggctacar          413
Gly His Arg Ile Cys Asp Leu
                   70
aaaccacagt cttcccagca attattacaa ttcttgaatt ccttggggat tttttactgc        473
cctttcaaag cacttaaktg tkrratctaa cgtkttccag tgtctgtctg aggtgactta        533
aaaaatcaga acaaaacttc tattatccag agtcatggga gagtacaccc tttccaggaa        593
taatgttttg ggaaacactg aaatgaaatc ttcccagtat tataaattgt gtatttaaaa        653
aaaagaaact tttctgaatg cctacctggc ggtgtatacc aggcagtgtg ccagtttaaa        713
aagatgaaaa agaataaaaa cttttgagga aaaaaaaaaa a                            754
```

```
<210> 298
<211> 629
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 66..497

<221> sig_peptide
<222> 66..239
<223> Von Heijne matrix
      score 5.40000009536743
      seq QLLDSVLWLGALG/LT

<221> polyA_signal
<222> 594..599

<221> polyA_site
<222> 618..629
```

```
<400> 298
aactcccaga atgctgacca aagtgggagg agcactaggt cttcccgtca cctccacctc          60
tctcc atg acc cgg ctc tgc tta ccc aga ccc gaa gca cgt gag gat ccg        110
      Met Thr Arg Leu Cys Leu Pro Arg Pro Glu Ala Arg Glu Asp Pro
             -55                        -50                       -45
atc cca gtt cct cca agg ggc ctg ggt gct ggg gag ggg tca ggt agt        158
Ile Pro Val Pro Pro Arg Gly Leu Gly Ala Gly Glu Gly Ser Gly Ser
             -40                       -35                  -30
cca gtg cgt cca cct gta tcc acc tgg ggc cct agc tgg gcc cag ctc         206
Pro Val Arg Pro Pro Val Ser Thr Trp Gly Pro Ser Trp Ala Gln Leu
```

```
        -25                    -20                    -15
ctg gac agt gtc cta tgg ctg ggg gca cta gga ctg aca atc cag gca      254
Leu Asp Ser Val Leu Trp Leu Gly Ala Leu Gly Leu Thr Ile Gln Ala
        -10                    -5                    1          5
gtc ttt tcc acc act ggc cca gcc ctg ctg ctg ctt ctg gtc agc ttc      302
Val Phe Ser Thr Thr Gly Pro Ala Leu Leu Leu Leu Leu Val Ser Phe
                10                    15                    20
ctc acc ttt gac ctg ctc cat agg ccc gca gtc aca ctc tgc cac agc      350
Leu Thr Phe Asp Leu Leu His Arg Pro Ala Val Thr Leu Cys His Ser
                25                    30                    35
gca aac ttc tca cca ggg gcc aga gtc agg ggg ccg gtg aag gtc ctg      398
Ala Asn Phe Ser Pro Gly Ala Arg Val Arg Gly Pro Val Lys Val Leu
                40                    45                    50
gac agc agg agg ctc tac tcc tgc aaa tgg gta cag tct cag gac aac      446
Asp Ser Arg Arg Leu Tyr Ser Cys Lys Trp Val Gln Ser Gln Asp Asn
        55                    60                    65
tta gcc tcc agg aag cac tgc tgc tgc tgc tca tgg ggc tgg gcc cgc      494
Leu Ala Ser Arg Lys His Cys Cys Cys Cys Ser Trp Gly Trp Ala Arg
70                    75                    80                    85
tcc tgaaaacctg tggcatgccc ttgwaccctg cttggcctgg ctttctgcct           547
Ser
ccatccttgg gcctgakanc ccctccccac aactcagtgt ccttcaaata tacaatgacc    607
acccttcttc aaaaaaaaaa aa                                            629
```

```
<210> 299
<211> 765
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 49..411

<221> sig_peptide
<222> 49..96
<223> Von Heijne matrix
      score 10.1000003814697
      seq LVLTLCTLPLAVA/SA

<221> polyA_signal
<222> 732..737

<221> polyA_site
<222> 750..763

<400> 299
aaagatccct gcagcccggc aggagagaag gctgagcctt ctggcgtc atg gag agg    57
                                                       Met Glu Arg
                                                               -15
ctc gtc cta acc ctg tgc acc ctc ccg ctg gct gtg gcg tct gct ggc     105
Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala Ser Ala Gly
                -10                    -5                    1
tgc gcc acg acg cca gct cgc aac ctg agc tgc tac cag tgc ttc aag     153
Cys Ala Thr Thr Pro Ala Arg Asn Leu Ser Cys Tyr Gln Cys Phe Lys
        5                    10                    15
gtc agc agc tgg acg gag tgc ccg ccc acc tgg tgc agc ccg ctg gac     201
Val Ser Ser Trp Thr Glu Cys Pro Pro Thr Trp Cys Ser Pro Leu Asp
20                    25                    30                    35
caa gtc tgc atc tcc aac gag gtg gtc gtc tct ttt agt gag tcy ccc     249
Gln Val Cys Ile Ser Asn Glu Val Val Val Ser Phe Ser Glu Ser Pro
```

```
                40                    45                    50
ccg ggc aga ggg cas gtg cca bgt gcc ggg gaa kgg ccg gtg ccc ccg      297
Pro Gly Arg Gly Xaa Val Pro Xaa Ala Gly Glu Xaa Pro Val Pro Pro
            55                    60                    65
cct ctc wkc gac tta bct atg act cct cgg ckc ycc agg gcc tgg ggc      345
Pro Leu Xaa Asp Leu Xaa Met Thr Pro Arg Xaa Xaa Arg Ala Trp Gly
            70                    75                    80
cck gtg ggt ccd aaa gtg cct cct gct gtc tct ccc gcg ctg ggc tcg      393
Pro Val Gly Pro Lys Val Pro Pro Ala Val Ser Pro Ala Leu Gly Ser
            85                    90                    95
ggc gag cat ccs rva btg tgaatkkkga ctttttctc ckccatttga              441
Gly Glu His Pro Xaa Xaa
100                    105
agtgtcacta ggaactgtca gcaggacaaa ggctctgatg tcactgaatt tacaaaraca    501
gcaggaacrs ackggtgggg atgggcagct gttcrargcr atgggtkatc tgcccttcct    561
ggcacagcac artacacctg ccatacaacc carcatcagg cakgctgcac tggaatcgat    621
acagtgtatg acaatgtcat atagtataac acaacataat gaatataacg tgtatattgc    681
aacttaatat aatacgatgt aatataatgc tacataatac aacataatat aataaaatag    741
aatgcaacac aaaaaaaaaa aacc                                           765
```

```
<210> 300
<211> 623
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 49..534

<221> sig_peptide
<222> 49..96
<223> Von Heijne matrix
      score 10.1000003814697
      seq LVLTLCTLPLAVA/SA

<221> polyA_signal
<222> 593..598

<221> polyA_site
<222> 612..623

<400> 300
aaagatccct gcagcccggc aggagagaag gctgagcctt ctggcgtc atg gag agg    57
                                                      Met Glu Arg
                                                              -15
ctc gtc cta acc ctg tgc acc ctc ccg ctg gct gtg gcg tct gct ggc    105
Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala Ser Ala Gly
            -10                   -5                    1
tgc gcc acg acg cca gct cgc aac ctg agc tgc tac cag tgc ttc aag    153
Cys Ala Thr Thr Pro Ala Arg Asn Leu Ser Cys Tyr Gln Cys Phe Lys
5                    10                    15
gtc agc agc tgg acg gag tgc ccg ccc acc tgg tgc agc ccg ctg gac    201
Val Ser Ser Trp Thr Glu Cys Pro Pro Thr Trp Cys Ser Pro Leu Asp
20                    25                    30                    35
caa gtc tgc atc tcc aac gag gtg gtc gtc tct ttt aaa tgg agt gta    249
Gln Val Cys Ile Ser Asn Glu Val Val Val Ser Phe Lys Trp Ser Val
            40                    45                    50
cgc gtc ctg ctc agc aaa cgc tgt gct ccc aga tgt ccc aac gac aac    297
Arg Val Leu Leu Ser Lys Arg Cys Ala Pro Arg Cys Pro Asn Asp Asn
            55                    60                    65
```

```
atg aak ttc gaa tgg tcg ccg gcc ccc atg gtg caa ggc gtg atc acc      345
Met Xaa Phe Glu Trp Ser Pro Ala Pro Met Val Gln Gly Val Ile Thr
        70                      75                  80
agg cgc tgc tgt tcc tgg gct ctc tgc aac agg gca ctg acc cca cag      393
Arg Arg Cys Cys Ser Trp Ala Leu Cys Asn Arg Ala Leu Thr Pro Gln
        85                      90                  95
gag ggg cgc tgg gcc ctg cra ggg ggg ctc ctc cag gac cct tcg          441
Glu Gly Arg Trp Ala Leu Xaa Gly Gly Leu Leu Gln Asp Pro Ser
100                     105                 110                 115
agg ggc ara aaa acc tgg gtg cgg cca cag ctg ggg ctc cca ctc tgc      489
Arg Gly Xaa Lys Thr Trp Val Arg Pro Gln Leu Gly Leu Pro Leu Cys
                120                     125                 130
ctt ccc awt tcc aac ccc ctc tgc cca rgg gaa acc cag gaa gga          534
Leu Pro Xaa Ser Asn Pro Leu Cys Pro Xaa Glu Thr Gln Glu Gly
                135                     140                 145
taacactgtg ggtgcccca cctgtgcatt gggaccacra cttcaccctc ttggaracaa     594
taaactctca tgcccccaaa aaaaaaaa                                        623
```

<210> 301
<211> 571
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 86..415

<221> sig_peptide
<222> 86..145
<223> Von Heijne matrix
       score 9.80000019073486
       seq FTIGLTLLLGXQA/MP

<221> polyA_signal
<222> 540..545

<221> polyA_site
<222> 560..571

<400> 301
```
aaaaactcac ccagtgagtg tgagcattta agaagcatcc tctgccaaga ccaaaaggaa      60
agaagaaaaa bggccaaaag ccaaa atg ara ctg atg gta ctt gtt ttc acc      112
                             Met Xaa Leu Met Val Leu Val Phe Thr
                             -20                 -15
att ggg cta act ttg ctg cta gga rtt caa gcc atg cct gca aat cgc      160
Ile Gly Leu Thr Leu Leu Leu Gly Xaa Gln Ala Met Pro Ala Asn Arg
    -10                 -5                  1            -      5
ctc tct tgc tac aga aag ata cta aaa gat cac aac tgt cac aac ctt      208
Leu Ser Cys Tyr Arg Lys Ile Leu Lys Asp His Asn Cys His Asn Leu
                10                  15                  20
ccg gaa gga gta gct gac ctg aca cag att gat gtc aat gtc cag gat      256
Pro Glu Gly Val Ala Asp Leu Thr Gln Ile Asp Val Asn Val Gln Asp
            25                  30                  35
cat ttc tgg gat ggg aag gga tgt gag atg atc tgt tac tgc aac ttc      304
His Phe Trp Asp Gly Lys Gly Cys Glu Met Ile Cys Tyr Cys Asn Phe
        40                      45                  50
aag cga att gct ctg ctg ccc aaa aga cgt ttt ctt tgg acc aaa gat      352
Lys Arg Ile Ala Leu Leu Pro Lys Arg Arg Phe Leu Trp Thr Lys Asp
        55                      60                  65
ctc ttt cgt gat tcc ttg caa caa tca atg aga atc ttc atg tat tct      400
```

```
Leu Phe Arg Asp Ser Leu Gln Gln Ser Met Arg Ile Phe Met Tyr Ser
70              75              80              85
ggc gaa cac cat tcc tgatttccca caaactgcac tacatcagta taactgcatt      455
Gly Glu His His Ser
             90
tctagtttct atatagtgca atagagcata gattctataa attcttactt gtctaagaaa    515
gtaaatctgt gttaaacaag tagtaataaa agttaattca atccaaaaaa aaaaaa        571
```

```
<210> 302
<211> 612
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 56..268

<221> sig_peptide
<222> 56..100
<223> Von Heijne matrix
      score 4.59999990463257
      seq LLTHNLLSSHVRG/VG

<221> polyA_signal
<222> 584..589

<221> polyA_site
<222> 601..612

<400> 302
ctaatcgaaa agggggattt tccggttccg gcctggcgag agtttgtgcg gcgac atg      58
                                                             Met
                                                             -15
aaa ctg ctt acc cac aat ctg ctg agc tcg cat gtg cgg ggg gtg ggg     106
Lys Leu Leu Thr His Asn Leu Leu Ser Ser His Val Arg Gly Val Gly
            -10             -5              1
tcc cgt ggc ttc ccc ctg cgc ctc cag gcc acc gag gtc cgt atc tgc     154
Ser Arg Gly Phe Pro Leu Arg Leu Gln Ala Thr Glu Val Arg Ile Cys
        5               10              15
cct gtg gaa ttc aac ccc aac ttc gtg gcg cgt atg ata cct aaa gtg     202
Pro Val Glu Phe Asn Pro Asn Phe Val Ala Arg Met Ile Pro Lys Val
    20              25              30
gag tgg tcg gcg ttc ctg gag gcg rmc gat aac ttg cgt ctg atc cag     250
Glu Trp Ser Ala Phe Leu Glu Ala Xaa Asp Asn Leu Arg Leu Ile Gln
35              40              45              50
gtg ccg aga agg gcc ggt tgagggatat gaggagaatg aggagtttct             298
Val Pro Arg Arg Ala Gly
             55
gaggaccatg caccacctgc tgctggaggt ggamstgaka gagggcaccc tgcagtgccc    358
ggaatctgga cgtatgttcc ccatcagccg cgggatcccc aacatgctgc tgagtgaaga    418
ggaaactgag agttgattgt gccaggcgcc agtttttctt gttatgactg tgtattttg     478
ttgatctata ccctgttcc gaattctgcc gtgtgtatcc ccaacccttg acccaatgac     538
accaaacaca gtgttttga gctcggtatt atatatttt ttctcattaa aggtttaaaa     598
ccaaaaaaaa aaaa                                                      612
```

```
<210> 303
<211> 539
<212> DNA
```

<213> Homo sapiens

<220>
<221> CDS
<222> 32..328

<221> sig_peptide
<222> 32..103
<223> Von Heijne matrix
      score 4.59999990463257
      seq FFIFCSLNTLLLG/GV

<221> polyA_signal
<222> 508..513

<221> polyA_site
<222> 528..539

<400> 303
aacaactatc ctgcctgctg cttgctgcac c atg aag tct gcc aag ctg gga   52
                                     Met Lys Ser Ala Lys Leu Gly
                                                             -20
ttt ctt cta aga ttc ttc atc ttc tgc tca ttg aat acc ctg tta ttg   100
Phe Leu Leu Arg Phe Phe Ile Phe Cys Ser Leu Asn Thr Leu Leu Leu
        -15   -10   -5
ggt ggt gtt aat aaa att gcg gag aag ata tgt gga gac ctc aaa gat   148
Gly Gly Val Asn Lys Ile Ala Glu Lys Ile Cys Gly Asp Leu Lys Asp
  1   5   10   15
ccc tgc aaa ttg gac atg aat ttt gga agc tgc tat gaa gtt cac ttt   196
Pro Cys Lys Leu Asp Met Asn Phe Gly Ser Cys Tyr Glu Val His Phe
                  20   25   30
aga tat ttc tac aac aga acc tcc aaa aga tgt gaa act ttt gtc ttc   244
Arg Tyr Phe Tyr Asn Arg Thr Ser Lys Arg Cys Glu Thr Phe Val Phe
              35   40   45
tcc agc tgt aat ggc aac ctt aac aac ttc aag ctt aaa ata gaa cgt   292
Ser Ser Cys Asn Gly Asn Leu Asn Asn Phe Lys Leu Lys Ile Glu Arg
          50   55   60
gaa gta kcc tgt gtt gca aaa tac aaa cca ccg agg tgagaggatg   338
Glu Val Xaa Cys Val Ala Lys Tyr Lys Pro Pro Arg
    65   70   75
tgaactcatg aagttgtctg ctgcaccatc cgaaataaag acacaagaaa attcaractg   398
atttwgaaat ctttgttwta tttccmymak ggcgwktaag cttccatatg tttgctattt   458
tcctgaccct agttttgtct ttcctggaaa ttaactgtat gakcattasa atgaaagagt   518
ctttctgtca aaaaaaaaaa a   539

<210> 304
<211> 964
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 21..527

<221> sig_peptide
<222> 21..95
<223> Von Heijne matrix
      score 8.5
      seq LKVLLLPLAPAAA/QD

<221> polyA_signal
<222> 921..926

<221> polyA_site
<222> 953..963

<400> 304

```
agggcggatc ttctccggcc atg agg aag cca gcc gct ggc ttc ctt ccc tca          53
                      Met Arg Lys Pro Ala Ala Gly Phe Leu Pro Ser
                      -25               -20               -15

ctc ctg aag gtg ctg ctc ctg cct ctg gca cct gcc gca gcc cag gat            101
Leu Leu Lys Val Leu Leu Leu Pro Leu Ala Pro Ala Ala Ala Gln Asp
            -10               -5                1

tcg act cag gcc tcc act cca ggc agc cct ctc tct cct acc gaa tac            149
Ser Thr Gln Ala Ser Thr Pro Gly Ser Pro Leu Ser Pro Thr Glu Tyr
        5                 10                15

caa cgc ttc ttc gca ctg ctg act cca acc tgg aag gca gar act acc            197
Gln Arg Phe Phe Ala Leu Leu Thr Pro Thr Trp Lys Ala Glu Thr Thr
        20                25                30

tgc cgt ctc cgt gca acc cac ggc tgc cgg aat ccc aca ctc gtc cag           245
Cys Arg Leu Arg Ala Thr His Gly Cys Arg Asn Pro Thr Leu Val Gln
35                40                45                5C

ctg gac caa tat gaa aac cac ggc tta gtg ccc gat ggt gct gtc tgc           293
Leu Asp Gln Tyr Glu Asn His Gly Leu Val Pro Asp Gly Ala Val Cys
            55                60                65

tcc aac ctc cct tat gcc tcc tgg ttt gag tct ttc tgc cag ttc act          341
Ser Asn Leu Pro Tyr Ala Ser Trp Phe Glu Ser Phe Cys Gln Phe Thr
            70                75                80

cac tac cgt tgc tcc aac cac gtc tac tat gcc aag aga gtc ctg tgt          389
His Tyr Arg Cys Ser Asn His Val Tyr Tyr Ala Lys Arg Val Leu Cys
            85                90                95

tcc cag cca gtc tct att ctc tcw cct aac act ctc aag gag ata gaa          437
Ser Gln Pro Val Ser Ile Leu Ser Pro Asn Thr Leu Lys Glu Ile Glu
        100               105               110

sct tca gct gaa gtc tca ccc acc aca gat gac ctc ccc cat ctc acc          485
Xaa Ser Ala Glu Val Ser Pro Thr Thr Asp Asp Leu Pro His Leu Thr
115               120               125               130

cca ctt cac agt gac aga acg cca gac ctt cca gcc ctg gcc                   527
Pro Leu His Ser Asp Arg Thr Pro Asp Leu Pro Ala Leu Ala
            135               140

tgagaggctc agcaacaacg tggaagagct cctacaatcc tccttgtccc tgggaggcca        587
ggagcaagcg ccagagcaca agcaggagca aggagtggag cacaggcagg agccgacaca        647
agaacacaag caggaagagg ggcagaaaca ggaagagcaa gaagaggaac aggaagagga        707
gggaaagcag gaagaaggac aggggactaa ggagggacgg gaggctgtgt ctcagctgca        767
gacagactca gagcccaagt ttcactctga atctctatct tctaacccct cctcttttgc        827
tccccgggta cganaagtag agtctactcc tatgataatg gagaacatcc aggagctcat        887
tcgatcagcc caggaaatag atgaaatgaa tgaaatatat gatgagaact cctactggag        947
aaaccaaaaa aaaaaak                                                       964
```

<210> 305
<211> 684
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 147..647

<221> sig_peptide
<222> 147..374

<223> Von Heijne matrix
      score 3.5
      seq LASASELPLGSRP/AP

<221> polyA_site
<222> 668..681

<400> 305
aacttcctgt gagcccggcg gtgacaacgg caacatggcc cgtgaacgga gctgaagtcg   60
acgacttctc ctrgrarmcc ccgactgagg cggagacgaa ggtgctgcag cgcgcgacggg   120
agcggcaaga tcgcatctcc cggctc atg ggc gac tat ctg ctg cgc ggt tac   173
                              Met Gly Asp Tyr Leu Leu Arg Gly Tyr
                                  -75   -70
cgc atg ctg ggc gag acg tgt gcg gac tgc ggg acg atc ctc ctc caa   221
Arg Met Leu Gly Glu Thr Cys Ala Asp Cys Gly Thr Ile Leu Leu Gln
        -65   -60   -55
gac aaa cag cgg aaa atc tac tgc gtg gct tgt cag gaa ctc gac tca   269
Asp Lys Gln Arg Lys Ile Tyr Cys Val Ala Cys Gln Glu Leu Asp Ser
    -50   -45   -40
gac gtg gat aaa gat aat ccc gct ctg aat gcc cag gct gcc ctc tcc   317
Asp Val Asp Lys Asp Asn Pro Ala Leu Asn Ala Gln Ala Ala Leu Ser
-35   -30   -25   -20
caa gct cgg gag cac cag ctg gcc tca gcc tca gag ctc ccc ctg ggc   365
Gln Ala Arg Glu His Gln Leu Ala Ser Ala Ser Glu Leu Pro Leu Gly
        -15   -10   -5
tct cga cct gcg ccc caa ccc cca gta cct cgt ccg gag cac tgt gag   413
Ser Arg Pro Ala Pro Gln Pro Pro Val Pro Arg Pro Glu His Cys Glu
        1   5   10
gga gct gca gca gga ctc aag gca gcc cag ggg cca cct gct cct gct   461
Gly Ala Ala Ala Gly Leu Lys Ala Ala Gln Gly Pro Pro Ala Pro Ala
    15   20   25
gtg cct cca aat aca rat gtc atg gcc tgc aca cag aca gcc ctc ttg   509
Val Pro Pro Asn Thr Xaa Val Met Ala Cys Thr Gln Thr Ala Leu Leu
30   35   40   45
caa aag ctg acc tgg gcc tct gct gaa ctg ggc tct anc acc tcc cyg   557
Gln Lys Leu Thr Trp Ala Ser Ala Glu Leu Gly Ser Xaa Thr Ser Xaa
        50   55   60
gga aaa mta gca tcc agc tgt gtg gcc tta tcc gcg cat gtg cgg agg   605
Gly Lys Xaa Ala Ser Ser Cys Val Ala Leu Ser Ala His Val Arg Arg
        65   70   75
ccc tgc gca gcc tgc agc agc tac agc act aag aga agc ccc   647
Pro Cys Ala Ala Cys Ser Ser Tyr Ser Thr Lys Arg Ser Pro
        80   85   90
tgagaaaaac ctctagaaaa acaaaaaaaa aaaaccc   684

<210> 306
<211> 693
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 262..471

<221> sig_peptide
<222> 262..306
<223> Von Heijne matrix
      score 3.5
      seq LCFLLPHHRLQEA/RQ

<221> polyA_signal
<222> 663..668

<221> polyA_site
<222> 682..693

<400> 306
atttcgcggc gctcgcbgma cyhsgwtgtt cagcaccttc ggtccggttg aggttgtcaa   60
gtcggmccaa acaggttgtt tctctgcagt ttccaacatg gcagggmsgt ttaatagaca   120
tggataagaa gtccactcac agaaatcctg aagatgccag ggctggcaaa tatgaaggta   180
aacacaaacg aaagaaaaga agaaagcaaa accaaaacca gcaccgatcc cgacatagat   240
cagtgacgtc ttttcttca g atg atc cta tgt ttc ctt ctt cct cat cat   291
                       Met Ile Leu Cys Phe Leu Leu Pro His His
                       -15   -10
cgt ctt cag gaa gcc aga cag att caa gta ttg aag atg ctt cca agg   339
Arg Leu Gln Glu Ala Arg Gln Ile Gln Val Leu Lys Met Leu Pro Arg
-5   1   5   10
gaa aaa tta aga aga aga gaa gag aga aaa caa ata aat ggg aaa aaa   387
Glu Lys Leu Arg Arg Arg Glu Glu Arg Lys Gln Ile Asn Gly Lys Lys
              15   20   25
raa agg aca aaa tat gaa aca cca aga aaa rga raa gga aaa aaa gga   435
Xaa Arg Thr Lys Tyr Glu Thr Pro Arg Lys Xaa Xaa Gly Lys Lys Gly
          30   35   40
gga aac mac cmc wtw tkt cmc ctt tcc aar agg gac tgaaactggg   481
Gly Asn Xaa Xaa Xaa Xaa Xaa Leu Ser Lys Arg Asp
          45   50   55
ctgacccttt tgatttccaa vctcascgtt ttggtgtaag gcggccaaar aaggatgcgg   541
ascccagcac tgtgaagcct acaaaaacat tgatgcgctg gcttggggat ttgaatttga   601
acatctttca cactaagttc agactcatga aaccaatctt cagatgctct gtaaaccaca   661
taataaagag tttggaaatt aaaaaaaaar aa   693

<210> 307
<211> 1656
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 74..1216

<221> sig_peptide
<222> 74..172
<223> Von Heijne matrix
      score 5.80000019073486
      seq XLCLGMALCPRQA/TR

<221> polyA_signal
<222> 1627..1632

<221> polyA_site
<222> 1640..1652

<400> 307
atctcttggc gtctcaacgt tcggatcagc agctttcttc cattctctct ctccacttct   60
tcagtgagca gcc atg agt tgg act gtg cct gtt gtg cgg gcc agc cag   109
            Met Ser Trp Thr Val Pro Val Val Arg Ala Ser Gln
                    -30   -25
aga gtg agc tcg gtg gga gcg aat ktc cta tgc ctg ggg atg gcc ctg   157
Arg Val Ser Ser Val Gly Ala Asn Xaa Leu Cys Leu Gly Met Ala Leu
    -20   -15   -10

```
tgt ccg cgt caa gca acg cgc atc ccg ctc aac ggc acc tgg ctc ttc    205
Cys Pro Arg Gln Ala Thr Arg Ile Pro Leu Asn Gly Thr Trp Leu Phe
-5               1               5                   10
acc ccc gtg agc aag atg gcg act gtg aar agt gag ctt att gag cgt    253
Thr Pro Val Ser Lys Met Ala Thr Val Lys Ser Glu Leu Ile Glu Arg
         15               20                  25
ttc act tcc gar aag ccc gtt cat cac agt aag gtc tcc atc ata gga    301
Phe Thr Ser Glu Lys Pro Val His His Ser Lys Val Ser Ile Ile Gly
         30               35                  40
act gga tcg gtg ggc atg gcc tgc gct atc agc atc tta tta aaa ggc    349
Thr Gly Ser Val Gly Met Ala Cys Ala Ile Ser Ile Leu Leu Lys Gly
     45               50                  55
ttg agt gat gaa ctt gcc ctt gtg gat ctt gat gaa rac aaa ctg aag    397
Leu Ser Asp Glu Leu Ala Leu Val Asp Leu Asp Glu Xaa Lys Leu Lys
60               65               70                  75
ggt gag acr atg gat ctt caa cat ggc agc cct ttc acg aaa atg cca    445
Gly Glu Thr Met Asp Leu Gln His Gly Ser Pro Phe Thr Lys Met Pro
             80               85                  90
aat att gtt tgt agc aaa rat tac ttt gtc aca gca aac tcc aac cta    493
Asn Ile Val Cys Ser Lys Xaa Tyr Phe Val Thr Ala Asn Ser Asn Leu
             95               100                 105
gtg att atc aca gca ggt gca cgc caa raa aag gga gaa acg cgc ctt    541
Val Ile Ile Thr Ala Gly Ala Arg Gln Xaa Lys Gly Glu Thr Arg Leu
         110               115                 120
aat tta stc cag cga aat gtg gcc atc ttc aag tta atg att tcc agt    589
Asn Leu Xaa Gln Arg Asn Val Ala Ile Phe Lys Leu Met Ile Ser Ser
     125               130                 135
att gtc cag tac agc ccc cac tgc aaa ctg att att gtt tcc aat cca    637
Ile Val Gln Tyr Ser Pro His Cys Lys Leu Ile Ile Val Ser Asn Pro
140               145               150                 155
gtg gat atc tta act tat gta gct tgg aag ttg agt gca ttt ccc aaa    685
Val Asp Ile Leu Thr Tyr Val Ala Trp Lys Leu Ser Ala Phe Pro Lys
             160               165                 170
aac cgt att att gga agc ggc tgt aat ctg ata mhg gct cgt ttt cgt    733
Asn Arg Ile Ile Gly Ser Gly Cys Asn Leu Ile Xaa Ala Arg Phe Arg
             175               180                 185
ttc ttg att gga caa aag ctt ggt atc cat tct gaa agc tgc cat gga    781
Phe Leu Ile Gly Gln Lys Leu Gly Ile His Ser Glu Ser Cys His Gly
         190               195                 200
tgg atc ctc gga gag cat gga gac tca agt gtt cct gtg tgg agt gga    829
Trp Ile Leu Gly Glu His Gly Asp Ser Ser Val Pro Val Trp Ser Gly
     205               210                 215
gtg aac ata gct ggt gtc cct ttg aag gat ctg aac tct gat ata gga    877
Val Asn Ile Ala Gly Val Pro Leu Lys Asp Leu Asn Ser Asp Ile Gly
220               225               230                 235
act gat aaa gat cct gag caa tgg aaa aat gtc cac aaa gaa gtg act    925
Thr Asp Lys Asp Pro Glu Gln Trp Lys Asn Val His Lys Glu Val Thr
             240               245                 250
gca act gcc tat gag att att aaa atg aaa ggt tat act tct tgg gcc    973
Ala Thr Ala Tyr Glu Ile Ile Lys Met Lys Gly Tyr Thr Ser Trp Ala
         255               260                 265
att ggc cta tct gtg gcc gat tta aca gaa agt att ttg aag aat ctt    1021
Ile Gly Leu Ser Val Ala Asp Leu Thr Glu Ser Ile Leu Lys Asn Leu
         270               275                 280
agg aga ata cat cca gtt tcc acc ata act aag ggc ctc tat gga ata    1069
Arg Arg Ile His Pro Val Ser Thr Ile Thr Lys Gly Leu Tyr Gly Ile
285               290                 295
rat gaa gaa gta ttc ctc agt att cct tgt atc ctg gga gag aac ggt    1117
Xaa Glu Glu Val Phe Leu Ser Ile Pro Cys Ile Leu Gly Glu Asn Gly
300               305                 310                 315
att acc aac ctt ata aag ata aag ctg acc cct gaa gaa gag gcc cat    1165
Ile Thr Asn Leu Ile Lys Ile Lys Leu Thr Pro Glu Glu Glu Ala His
```

```
                    320                    325                    330
          ctg aaa aaa agt gca aaa aca ctc tgg gaa att cag aat aag ctt aag     1213
          Leu Lys Lys Ser Ala Lys Thr Leu Trp Glu Ile Gln Asn Lys Leu Lys
                    335                    340                    345
          ctt taaagttgcc taaaactacc attccgaaat tattgaagag atcatagata          1266
          Leu
          caggattata taacgaaatt ttgaataaac ttgaattcct aaaagatgga aacaggaaag   1326
          taggtagagt gattttccta tttatttagt cctccagctc ttttattgag catccacgtg   1386
          ctggacgata cttatttaca attcckaagt attttggta cctctgatgt agcagcactt     1446
          gccatgttat atatatgtag ttgrmatttg gttcccaaaa agtaggatgt aggtatttat    1506
          tgtgttctag aaattccgac tcttttcatt agatatatgc tatttctttc attcttgctg    1566
          gtttatacct atgttcattt atatgctgta aaaaagtagt agcttcttct acaatgtaaa    1626
          aataaatgta catacaaaaa aaaaaamcmc                                    1656
```

<210> 308
<211> 517
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 48..164

<221> sig_peptide
<222> 48..89
<223> Von Heijne matrix
      score 4
      seq YYMVCLFFRLIFS/EH

<221> polyA_signal
<222> 482..487

<221> polyA_site
<222> 505..517

<400> 308
```
          aggagatagc ctcgtagaaa tgacaaccac aatgttaata ctaacat atg tat tac      56
                                                                  Met Tyr Tyr
          atg gtt tgt ttg ttc ttt cgc tta ata ttt tca gag cac cta cct att     104
          Met Val Cys Leu Phe Phe Arg Leu Ile Phe Ser Glu His Leu Pro Ile
              -10                    -5                     1                5
          ata ggc act gtc act tct cac aaa act ggg aca cta act gtt tat cca     152
          Ile Gly Thr Val Thr Ser His Lys Thr Gly Thr Leu Thr Val Tyr Pro
                            10                     15                     20
          aca tct gct ggc taaataaaga catgatcttc accttttggg attgttaatt        204
          Thr Ser Ala Gly
                        25
          taaaatggtt ccataagagc aatgcaaaga cagagatatt tggcagcact gcagctggtg   264
          atttatatgg ctcttcacaa ggtgttattt tggggtatca aggtatggat gcttaaatca   324
          gctgcaggaa gtaagaaaga agaaaaaagg agtgataaag ataaaaaaaa atcaaccttg   384
          gtccttccac caaaacccat taatttccat atcatcatct gcataararg gaaaattcct   444
          acwtgaccag gttactgcaa ggatktkaat tttgaatatt aaaatattat mcmcaattgg   504
          aaaaaaaaaa aaa                                                      517
```

<210> 309
<211> 405
<212> DNA
<213> Homo sapiens

```
<220>
<221> CDS
<222> 185..334

<221> sig_peptide
<222> 185..295
<223> Von Heijne matrix
      score 5.90000009536743
      seq LSYASSALSPCLT/AP

<221> polyA_signal
<222> 355..360

<221> polyA_site
<222> 392..405

<400> 309
atcaccttct tctccatcct tstctgggcc agtccccarc ccagtccctc tcctgacctg      60
cccagcccaa gtcagccttc agcacgcgct tttctgcaca cagatattcc aggcctacct     120
ggcattccag gacctccgma atgatgctcc agtcccttac aagcgcttcc tggatgaggg     180
tggc atg gtg ctg acc acc ctc ccc ttg ccc tct gcc aac agc cct gtg     229
     Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val
          -35                -30                -25
aac atg ccc acc act ggc ccc aac agc ctg agt tat gct agc tct gcc     277
Asn Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala
          -20                -15                -10
ctg tcc ccc tgt ctg acc gct cca aag tcc ccc cga ctt gct atg atg     325
Leu Ser Pro Cys Leu Thr Ala Pro Lys Ser Pro Arg Leu Ala Met Met
     -5                 1              5                  10
cct gac aac taaatatcct tatccaaatc aataaarwra raatcctccc            374
Pro Asp Asn
tccaraaggg tttctaaaaa caaaaaaaaa a                                405
```

```
<210> 310
<211> 1087
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 195..347

<221> sig_peptide
<222> 195..272
<223> Von Heijne matrix
      score 7.09999990463257
      seq LASLQWSLTLAWC/GS

<221> polyA_signal
<222> 1037..1042

<221> polyA_site
<222> 1071..1082

<400> 310
aaagtgtaga acacggacct ctgagttatg ctcttgagag gtgccaaagc tgggctgttt      60
acctacctta tccacagagc tctgaaagtc aagccagaaa ggaaggattc caaattcttg     120
gaattttatc tagaaaagaa gactaagcag cttttgttct tctgtgaccc agttgctggc     180
ccaagacatg gaca atg acc ccc tgg tgt ttg gcg tgt ctg ggg agg agg     230
```

```
                  Met Thr Pro Trp Cys Leu Ala Cys Leu Gly Arg Arg
                      -25                 -20                 -15
cct ctc gct tct ttg cag tgg agc ctg aca ctg gcg tgg tgt ggc tcc    278
Pro Leu Ala Ser Leu Gln Trp Ser Leu Thr Leu Ala Trp Cys Gly Ser
            -10                 -5                  1
ggc agc cac tgg aca gag aga cca akt cag akt tca ccg tgg akt tct    326
Gly Ser His Trp Thr Glu Arg Pro Xaa Gln Xaa Ser Pro Trp Xaa Ser
        5                   10                  15
ctg tca gcg acc acc agg ggg tgatcacacg gaaggtgaac atccaggtcg       377
Leu Ser Ala Thr Thr Arg Gly
    20                  25
gggatgtgaa tgacaacgcg cccacatttc acaatcagcc ctacagcgtc cgcatccctg   437
araatacacc agtgggggac cccatcttca tcgtgaatgc cacagacccc gacttggggg   497
cagggggcag cgtcctctac tccttccagc ccccctccca attcttcgtc attgacgcg    557
cccgcggtat cktcacagtg atccgggagc tggactacga taccacrcmg gcctaccagc   617
tcwcggtcwa cgccacagat caagacaara ccaggcctct gtccaccstg gccaacttgg   677
ccatcatcat cacagatgtc caggacatgg accccatctt catcaacctg ccttacagca   737
ccaacatcta cgagcattct cctccgggca cgacggtgcg catcatcacc gccatagacc   797
aggataaagg acgtccccgg ggcattggct acaccatcgt ttcagggcat ctgtgtttac   857
aagaacccaa gatctctcag gagctcagga aaaggggctt gctgtgaggc tcagggttcc   917
catggacatt ctgagctgac cctcctcagc attggatctc ctggctcagg aactaggaac   977
gaagcttgga tgttttctcc tttcctacag catctgtatt catttcctat agttgccata  1037
ataaaatgcc actaacttag tggcttaaaa accaaaaaaa aaaaacccctt            1087
```

<210> 311
<211> 916
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 90..815

<221> sig_peptide
<222> 90..179
<223> Von Heijne matrix
      score 13.1999998092651
      seq LLLLSTLVIPSAA/AP

<221> polyA_signal
<222> 883..888

<221> polyA_site
<222> 905..916

<400> 311
```
aaaacagtac gtgggcggcc ggaatccggg agtccggtga cccgggctgt ggtctagcat     60
aaaggcggag ccagaagaag gggcgggggt atg gga gaa gcc tcc cca cct gcc    113
                                  Met Gly Glu Ala Ser Pro Pro Ala
                                  -30                 -25
ccc gca agg cgg cat ctg ctg gtc ctg ctg ctg ctc ctc tct acc ctg    161
Pro Ala Arg Arg His Leu Leu Val Leu Leu Leu Leu Leu Ser Thr Leu
        -20                 -15                 -10
gtg atc ccc tcc gct gca gct cct atc cat gat gct gac gcc caa gag    209
Val Ile Pro Ser Ala Ala Ala Pro Ile His Asp Ala Asp Ala Gln Glu
    -5                  1                   5                   10
agc tcc ttg ggt ctc aca ggc ctc cag agc cta ctc caa ggc ttc agc    257
Ser Ser Leu Gly Leu Thr Gly Leu Gln Ser Leu Leu Gln Gly Phe Ser
                15                  20                  25
cga ctt ttc ctg aaa ggt aac ctg ctt cgg ggc aca gac agc tta ttc    305
```

```
        Arg Leu Phe Leu Lys Gly Asn Leu Leu Arg Gly Ile Asp Ser Leu Phe
                     30              35              40
tct gcc ccc atg gac ttc cgg ggc ctc cct ggg aac tac cac aaa gag      353
Ser Ala Pro Met Asp Phe Arg Gly Leu Pro Gly Asn Tyr His Lys Glu
             45              50              55
gag aac cag gag cac cag ctg ggg aac aac acc ctc tcc agc cac ctc      401
Glu Asn Gln Glu His Gln Leu Gly Asn Asn Thr Leu Ser Ser His Leu
     60              65              70
cag atc gac aag atg acc gac aac aag aca gga gag gtg ctg atc tcc      449
Gln Ile Asp Lys Met Thr Asp Asn Lys Thr Gly Glu Val Leu Ile Ser
75              80              85              90
gag aat gtg gtg gca tcc att caa cca vcg gag ggg anc ttc gag ggt      497
Glu Asn Val Val Ala Ser Ile Gln Pro Xaa Glu Gly Xaa Phe Glu Gly
                 95              100             105
gat ttg aag gth ccc agg atg gag gar aag gag gcc ctg gta ccc mtc      545
Asp Leu Lys Val Pro Arg Met Glu Glu Lys Glu Ala Leu Val Pro Xaa
             110             115             120
car aay gcc acg gac agc ttc cac aca gaa ctc cat ccc cgg gtg gcc      593
Gln Lys Ala Thr Asp Ser Phe His Thr Glu Leu His Pro Arg Val Ala
             125             130             135
ttc tgg atc att aag ctg cca cgg cgg agg tcc cac cag gat gcc ctg      641
Phe Trp Ile Ile Lys Leu Pro Arg Arg Arg Ser His Gln Asp Ala Leu
     140             145             150
gag ggc ggc cac tgg ctc anc gar aag cga cac cgc ctg cag gcc atc      689
Glu Gly Gly His Trp Leu Xaa Glu Lys Arg His Arg Leu Gln Ala Ile
155             160             165             170
cgg gat gga ctc cgc aag ggg acc cac aag gac rtc cta daa rag ggg      737
Arg Asp Gly Leu Arg Lys Gly Thr His Lys Asp Xaa Leu Xaa Xaa Gly
                 175             180             185
acc gar agc tcc tcc cac tcc agg ctg tcc ccc cga aar amm cac tta      785
Thr Glu Ser Ser Ser His Ser Arg Leu Ser Pro Arg Lys Xaa His Leu
             190             195             200
ctg tac atc ctc arg ccc tct cgg cag ctg targggtggg daccggggar      835
Leu Tyr Ile Leu Xaa Pro Ser Arg Gln Leu
             205             210
macctgcctg tagcccccat caraccctgc cccaagcacc atatggaaat aaagttcttt    895
cttacatcca aaaaaaaaaa a                                              916
```


```
<210> 312
<211> 583
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 52..513

<221> sig_peptide
<222> 52..231
<223> Von Heijne matrix
      score 4
      seq LVRRTLLVAALRA/WM

<221> polyA_signal
<222> 553..558

<221> polyA_site
<222> 572..583

<400> 312
```

```
aaggaaacag caaccagagg gagatgatca cctgaaccac tgctccaaac c atg ggc          57
                                                            Met Gly
                                                            -60
agt aaa tgc tgt aaa ggt ggt cca gat gaa gat gca gta gaa aga cag         105
Ser Lys Cys Cys Lys Gly Gly Pro Asp Glu Asp Ala Val Glu Arg Gln
        -55                 -50                 -45
agg cgg cag aag ttg ctt ctt gca caa ctg cat cac aga aaa agg gtg         153
Arg Arg Gln Lys Leu Leu Leu Ala Gln Leu His His Arg Lys Arg Val
        -40                 -35                 -30
aar gca gct ggg cag atc cag gcc tgg tgg cgt ggg gtc ctg gtg cgc         201
Lys Ala Ala Gly Gln Ile Gln Ala Trp Trp Arg Gly Val Leu Val Arg
    -25                 -20                 -15
agg acc ctg ctg gtt gct gcc ctc agg gcc tgg atg att cag tgc tgg         249
Arg Thr Leu Leu Val Ala Ala Leu Arg Ala Trp Met Ile Gln Cys Trp
-10                  -5                  1                   5
tgg agg acg ttg gtg cag aga cgg atc cgt cag cgg cgg cag gcc ctg         297
Trp Arg Thr Leu Val Gln Arg Arg Ile Arg Gln Arg Arg Gln Ala Leu
                10                  15                  20
ttr ggg gtc tac gtc atc cag gag cag gcg gcg gtc aag ctc cag tcc         345
Leu Gly Val Tyr Val Ile Gln Glu Gln Ala Ala Val Lys Leu Gln Ser
            25                  30                  35
tgc atc cgc atg tgg cag tgc cgg caa tgt tac cgc caa atg tgc aat         393
Cys Ile Arg Met Trp Gln Cys Arg Gln Cys Tyr Arg Gln Met Cys Asn
        40                  45                  50
gct ctc tgc ttg ttc cag gtc cca aaa agc agc ctt gcc ttc caa act         441
Ala Leu Cys Leu Phe Gln Val Pro Lys Ser Ser Leu Ala Phe Gln Thr
55                  60                  65                  70
gat ggc ttt tta cag gtc caa tat gca atc cct tca aag cag cca gag         489
Asp Gly Phe Leu Gln Val Gln Tyr Ala Ile Pro Ser Lys Gln Pro Glu
            75                  80                  85
ttc cac att gaa atc cta tca atc tgaaaggcct ggggcatgga gaacaggctg         543
Phe His Ile Glu Ile Leu Ser Ile
            90
cactacccta ataaatgtct gaccaggtaa aaaaaaaaa                               583
```

```
<210> 313
<211> 697
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 172..438

<221> sig_peptide
<222> 172..354
<223> Von Heijne matrix
      score 4.69999980926514
      seq LLPCNLHCSWLHS/SP

<221> polyA_signal
<222> 682..687

<221> polyA_site
<222> 685..697

<400> 313
agattggctg ggcagatggg ctgactggct gggcagatgg gtgggtgagt tccctctccc          60
cagagccatc ggccaggtac caaagctcag ctgtatggat tcccaacagg aggacctgcg         120
cttccctggg acccattgtt gtactggatt aacaagcgac ggcgctacgg c atg aat         177
```

```
                                                        Met Asn
                                                            -60
gca gcc atc aac acg ggc cct gcc cct gct gtc acc aag act gag act      225
Ala Ala Ile Asn Thr Gly Pro Ala Pro Ala Val Thr Lys Thr Glu Thr
        -55                 -50                 -45
gag gtc cag aat cca gat gtt ctg tgg gat ttg gac atc ccc gaa gcc      273
Glu Val Gln Asn Pro Asp Val Leu Trp Asp Leu Asp Ile Pro Glu Ala
        -40                 -35                 -30
agg agc cat gct gac caa gac agc aac ccc aag gcg gaa gcc ctg ctc      321
Arg Ser His Ala Asp Gln Asp Ser Asn Pro Lys Ala Glu Ala Leu Leu
        -25                 -20                 -15
ccc tgc aac ctg cac tgc agc tgg ctc cac agc agc ccc agg cca gat      369
Pro Cys Asn Leu His Cys Ser Trp Leu His Ser Ser Pro Arg Pro Asp
        -10                 -5                  1                 5
ccc cat tcc cac ttc cca tct ktc agg agg tgc cct ttg ccc cac cct      417
Pro His Ser His Phe Pro Ser Xaa Arg Arg Cys Pro Leu Pro His Pro
                    10                  15                  20
tgt gca acc tac ccc ccs kgc tgaaccactc tgtctcctat cctttggcca         468
Cys Ala Thr Tyr Pro Pro Xaa
                    25
cctgtcctga aaggaatgtt ctcttccatt ccctcctgaa tctggcccag gaagaccata    528
gcttcaatgy caagcctttt ccttcaaaac tgtagcctcc tctcactgaa ggtgggagct    588
gcaggaatca ggtgcagagt aggaaatgga actaacctca ggaaggtggt attgacagag    648
gtcaggaccc acctggatgt catgctatga aacattaaaa gaaaaaaaa                697
```

```
<210> 314
<211> 803
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 148..366

<221> sig_peptide
<222> 148..225
<223> Von Heijne matrix
      score 5.5
      seq LFTLLFLIMLVLK/LD

<221> polyA_signal
<222> 770..775

<221> polyA_site
<222> 792..803

<400> 314
aaatgggggg aaaagggcgg aaaaggacaa ggatccaaac tggcgaattt gctgatcttc    60
gcgtccctct ccgctttccg gccggcagcg ctgccagggt atattccttt tttccgatc     120
ctgcaacagc ctctttaaac tgtttaa atg aga atg tcc ttg gct cag aga gta    174
                              Met Arg Met Ser Leu Ala Gln Arg Val
                                  -25                 -20
cta ctc acc tgg ctt ttc aca cta ctc ttc ttg atc atg ttg gtg ttg      222
Leu Leu Thr Trp Leu Phe Thr Leu Leu Phe Leu Ile Met Leu Val Leu
        -15                 -10                 -5
aaa ctg gat gag aaa gca cct tgg aac tgg ttc ctc ata ttc att cca      270
Lys Leu Asp Glu Lys Ala Pro Trp Asn Trp Phe Leu Ile Phe Ile Pro
1                   5                   10                  15
gtc tgg ata ttt gat act atc ctt ctt gtc ctg ctg att gtg aaa atg      318
Val Trp Ile Phe Asp Thr Ile Leu Leu Val Leu Leu Ile Val Lys Met
```

```
                   20                      25                      30
gct ggg cgg tgt aag tct ggc ttt gac ctc gac atg gat cac aca ata        366
Ala Gly Arg Cys Lys Ser Gly Phe Asp Leu Asp Met Asp His Thr Ile
               35                      40                      45
taaaaaaaaa aacctggtac ctcattgcac tgtkacttaa attasccttc tgcctcgcac        426
tctgtgctaa actggaacag tttactacca tgaatctatc ctatgtcttc attcctttat        486
gggccttgct ggctggggct ttaacagaac tcggatataa tgtctttttt gtgaaagact        546
gacttctaag tacatcatct cctttctatt gctgttcaac aagttaccat taaagtgttc        606
tgaatctgtc aagcttcaag aataccagag aactgaggga aaataccaaa tgtagtttta        666
tactacttcc ataaaacagg attggtgaat cacggacttc tagtcaacct acagcttaat        726
tattcagcat ttgagttatt gaaatcctta ttatctctat gtaaataaag tttgttttgg        786
acctcaaaaa aaaaaaa                                                       803
```

```
<210> 315
<211> 823
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 175..336

<221> sig_peptide
<222> 175..276
<223> Von Heijne matrix
      score 3.70000004768372
      seq SVLNVGHLLFSSA/CS

<221> polyA_site
<222> 812..823

<400> 315
aaggcgcgcg cgaccggcgg ctctttggcg cggattaggg ggtctcggcg agggagtcat         60
caagctttgg tgtatgtgtt ggccggttct gaagtcttga agaagctctg ctgaggaaga        120
ccaaagcagc actcgttgcc aattagggaa tggaccgttt gggttccttt agca atg        177
                                                                  Met
atc cct ctg ata agc cac ctt gcc gag gct gct cct cct acc tca tgg        225
Ile Pro Leu Ile Ser His Leu Ala Glu Ala Ala Pro Pro Thr Ser Trp
            -30                     -25                     -20
agc ctt ata tca agt gtg ctg aat gtg ggc cac ctc ctt ttt tcc tct        273
Ser Leu Ile Ser Ser Val Leu Asn Val Gly His Leu Leu Phe Ser Ser
            -15                     -10                      -5
gct tgc agt gtt tca ctc gag gct ttg agt aca aga aac atc aaa gcg        321
Ala Cys Ser Val Ser Leu Glu Ala Leu Ser Thr Arg Asn Ile Lys Ala
 1               5                      10                      15
atc ata ctt atg aaa taatggcttc agattttcct gtccttgatc ccagctggac        376
Ile Ile Leu Met Lys
                20
tgctcaagaa raaatggccc ttttagaasc tgtgatggac tgtggctttg gaaattggca        436
ggatgtagcc aatcaaatgt gcaccaarac caaggaggag tgtgagaagc actatatgaa        496
gcatttcatc aataacccyc tgtttgcatc trscctgctg aacctgaaac aascagrgga        556
agcaaaaact gctgacacag ccattccatt tcactctaca ratgaccctc cccgacckac        616
ctttgactcc ttgctttctc gggacatggc cgggtacwtg ccmgctcgag cagatttcat        676
tgaggaattt gacaattatg cagaatggga cttgagagac attgattttg ttgaagatga        736
ctcggacatt ttacatgctc tgaagatggc tgtggtagat atctatcatt ccaggttaaa        796
ggagagacaa agacgaaaaa aaaaaaa                                           823
```

```
<210> 316
```

<211> 823
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 191..553

<221> sig_peptide
<222> 191..304
<223> Von Heijne matrix
      score 5.69999980926514
      seq LAFLSCLAFLVLD/TQ

<221> polyA_signal
<222> 766..771

<221> polyA_site
<222> 804..817

<400> 316
aactctgcag ggcctccaag gccaggcttc agggctggga ctcagtcctg aggcactggg   60
gagccatgag gggctgtggc agggaggggc agggtgtgga aagactcccc tggggccatg   120
gtggagatgt gctgaggtct tctccctgat cgtcttctcc tccctgctga ccgacggcta   180
ccagaackag atg gag tct ccg cag ctc cac tgc att ctc aac agc aac   229
              Met Glu Ser Pro Gln Leu His Cys Ile Leu Asn Ser Asn
                  -35   -30
agc gtg gcc tgc agc ttt gcc gtg gga gcc ggc ttc ctg gcc ttc ctc   277
Ser Val Ala Cys Ser Phe Ala Val Gly Ala Gly Phe Leu Ala Phe Leu
-25   -20   -15   -10
agc tgc ctg gcc ttc ctc gtc ctg gac aca cag gag acc cgc att gcc   325
Ser Cys Leu Ala Phe Leu Val Leu Asp Thr Gln Glu Thr Arg Ile Ala
                -5   1   5
ggc acc cgc ttc aag aca gcc ttc cag ctc ctg gac ttc atc ctg gct   373
Gly Thr Arg Phe Lys Thr Ala Phe Gln Leu Leu Asp Phe Ile Leu Ala
            10   15   20
gtt ctc tgg gca gtt gtc tgg ttc atg ggt ttc tgc ttc ctg gcc aac   421
Val Leu Trp Ala Val Val Trp Phe Met Gly Phe Cys Phe Leu Ala Asn
        25   30   35
caa tgg cag cat tcg ccg ccc aaa gar kkc ctc ctg ggg agc agc agt   469
Gln Trp Gln His Ser Pro Pro Lys Glu Xaa Leu Leu Gly Ser Ser Ser
40   45   50   55
gcc cag gca gcc atc ggc stt cac ctt ctt ctc cat cct tgt ctg gat   517
Ala Gln Ala Ala Ile Gly Xaa His Leu Leu Leu His Pro Cys Leu Asp
                60   65   70
att cca rgc cta cct ggc akk cca gga cct ccg aaa tgatgctcca   563
Ile Pro Xaa Leu Pro Gly Xaa Pro Gly Pro Pro Lys
                75   80
gtcccttacm arcgcttcct ggatgaaggt ggcatggtgs kkaacaccct cccccttgccc   623
tctgccaaca gcctgtgaac atgcccacca ctggccccaa cagcctgagt tatgctagct   683
ctgccctgtc cccctgtctg accgctcmaa agtcccccccg gcttgctatg atgcctgaca   743
actaaatatc cttatccaaa tcaataaaga gagaatcctc cctccagaag ggtttctaaa   803
aacaaaaaaa aaaahncctt   823

<210> 317
<211> 1112
<212> DNA
<213> Homo sapiens

<220>

<221> CDS
<222> 106..603

<221> sig_peptide
<222> 106..216
<223> Von Heijne matrix
    score 4.30000019073486
    seq LWEKLTLLSPGIA/VT

<221> polyA_site
<222> 1102..1112

<400> 317
agcgattgcg aatcctccgc tgaggtgatt tggatatccc tagaacgttg agggcacgag   60
tcgggtcctg agaccaggtc ctcagccagc agagccacgt tcctt atg agc acc gtg   117
                                                   Met Ser Thr Val
                                                                -35

ggt tta ttt cat ttt cct aca cca ctg acc cga ata tgc ccg gcc cca   165
Gly Leu Phe His Phe Pro Thr Pro Leu Thr Arg Ile Cys Pro Ala Pro
        -30   -25   -20

tgg gga ctc cgg ctt tgg gag aag ctg acg ttg tta tcc cca gga ata   213
Trp Gly Leu Arg Leu Trp Glu Lys Leu Thr Leu Leu Ser Pro Gly Ile
    -15   -10   -5

gct gtc act ccg gtc cag atg gca ggc aag aag gac tac cct gca ctg   261
Ala Val Thr Pro Val Gln Met Ala Gly Lys Lys Asp Tyr Pro Ala Leu
1   5   10   15

ctt tcc ttg gat gag aat gaa ctc gaa gag cag ttt gtg aaa gga cac   309
Leu Ser Leu Asp Glu Asn Glu Leu Glu Glu Gln Phe Val Lys Gly His
            20   25   30

ggt cca ggg ggc cag gca acc aac aaa acc agc aac tgc gtg gtg ctg   357
Gly Pro Gly Gly Gln Ala Thr Asn Lys Thr Ser Asn Cys Val Val Leu
        35   40   45

aar mac atc ccc tca ggc atc gtt gta aag tgc cat cag aca aga tca   405
Lys Xaa Ile Pro Ser Gly Ile Val Val Lys Cys His Gln Thr Arg Ser
        50   55   60

gtt gat cag aac aga aag cta gct cgg aaa atc cta caa gag aaa gta   453
Val Asp Gln Asn Arg Lys Leu Ala Arg Lys Ile Leu Gln Glu Lys Val
65   70   75

rat gtt ttc tac aat ggt gaa aac agt cct gtt cac aaa gaa aaa cga   501
Xaa Val Phe Tyr Asn Gly Glu Asn Ser Pro Val His Lys Glu Lys Arg
80   85   90   95

gaa gcg gcg aag aaa aaa car gaa agg aaa aaa aga gca aag gaa acc   549
Glu Ala Ala Lys Lys Lys Gln Glu Arg Lys Lys Arg Ala Lys Glu Thr
                100   105   110

ctg gaa aaa aag aas ctm ctt aaa raa ctg tgg gag tca agt aaa aag   597
Leu Glu Lys Lys Xaa Leu Leu Lys Xaa Leu Trp Glu Ser Ser Lys Lys
            115   120   125

gtc cac tgagaaaaga attagagatt ccaactgaca gaatctgcca gaagctccca   653
Val His
gggaataatg gtggcgagtt ccatcaccag cattattata gtgcttcaaa agaaatattt   713
ttgatgaact taaaagacaa caaatttatt taaatggtgc actaaactgt agtgaacaga   773
gacatgcacg attcaagaat aaaactcggc cgggcacggt ggacggtgcc tcacatctgt   833
aatcccagca ctttgggagg ccgaggcggg cggatcactt gaggtcagga gtttgagacc   893
agcctggcca acatggtgaa accccgtctc tactaaaaat acaaaaaatt agccaggcat   953
ggtggcgggc acctgtaatc ccagctactc gggaggccga ggcaggagaa ttgcgtgaac   1013
ctgggaggcg gaggttgcag tgagctgaga tcgcgccact gcactcaagc ctgggcaaca   1073
cctgggtgac agagcaagac cccatcycaa aaaaaaaaa   1112

<210> 318
<211> 1623

```
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 47..586

<221> sig_peptide
<222> 47..124
<223> Von Heijne matrix
      score 6.30000019073486
      seq GVGLVTLLGLAVG/SY

<221> polyA_signal
<222> 1583..1588

<221> polyA_site
<222> 1614..1623

<400> 318
aggatctgt cggcttgtca ggtggtggag gaaaaggcgc tccgtc atg ggg atc          55
                                                  Met Gly Ile
                                                      -25

cag acg agc ccc gtc ctg ctg gcc tcc ctg ggg gtg ggg ctg gtc act      103
Gln Thr Ser Pro Val Leu Leu Ala Ser Leu Gly Val Gly Leu Val Thr
        -20               -15                   -10

ctg ctc ggc ctg gct gtg ggc tcc tac ttg gtt cgg agg tcc cgc cgg      151
Leu Leu Gly Leu Ala Val Gly Ser Tyr Leu Val Arg Arg Ser Arg Arg
        -5                1               5

cct cag gtc act ctc ctg gac ccc aat gaa aag tac ctg cta cga ctg      199
Pro Gln Val Thr Leu Leu Asp Pro Asn Glu Lys Tyr Leu Leu Arg Leu
10                15                  20                  25

cta gac aag acg act gtg agc cac aac acc aag agg ttc cgc ttt gcc      247
Leu Asp Lys Thr Thr Val Ser His Asn Thr Lys Arg Phe Arg Phe Ala
            30                  35                  40

ctg ccc acc gcc cac cac act ctg ggg ctg cct gtg ggc aaa cat atc      295
Leu Pro Thr Ala His His Thr Leu Gly Leu Pro Val Gly Lys His Ile
            45                  50                  55

tac ctc tcc acm mga att gat ggc agc ctg gtc atc agg cca tac act      343
Tyr Leu Ser Thr Arg Ile Asp Gly Ser Leu Val Ile Arg Pro Tyr Thr
            60                  65                  70

cct gtc acc agt gat gag gat caa ggc tat gtg gat ctt gtc mtc aag      391
Pro Val Thr Ser Asp Glu Asp Gln Gly Tyr Val Asp Leu Val Xaa Lys
75.                     80                  85

gtc tac ctg aag ggt gtg cac ccc aaa ttt cct gag gga ggg aar atg      439
Val Tyr Leu Lys Gly Val His Pro Lys Phe Pro Glu Gly Gly Lys Met
90                  95                  100                 105

tct cak tac ctg gat asc ctg aaa gtt ggg gat btg gtg gaa ttt csg      487
Ser Xaa Tyr Leu Asp Xaa Leu Lys Val Gly Asp Xaa Val Glu Phe Xaa
            110                 115                 120

ggg cca agc ggg ttg ctc act tac act gga aaa ggg cat ttt aac att      535
Gly Pro Ser Gly Leu Leu Thr Tyr Thr Gly Lys Gly His Phe Asn Ile
            125                 130                 135

cag ccc aac aag aat ctc cac cag aac ccc gag tgg cga aga aac tgg      583
Gln Pro Asn Lys Asn Leu His Gln Asn Pro Glu Trp Arg Arg Asn Trp
            140                 145                 150

gaa tgattgccgg cgggacagga atcaccccaa tgctacagct gatccgggcc           636
Glu
atcctgaaag tccctgaaga tccaacccag tgctttctgc tttttgccaa ccagacagaa   696
aaggatatca tcttgcggga ggacttagag gaactgcagg cccgctatcc caatcgcttt   756
aagctctggt tcactctgga tcatcccca aaagrttggg cctacagcaa gggctttgtg    816
actgccgacw tgatccggga acacctgccc gctccagggg atgatgtgct ggtactgctt    876
```

```
tgtgggccmc ccccaatggt gcagctggcc tgccatccca acttggacaa actgggctac      936
tcacaaaaga tgcgattcac ctactgagca tcctccagct tccctggtgc tgttcgctgc      996
agttgttccc catcagtact caagcactak aagccttagr ktcctktcct cagagtttca     1056
ggtttttttca gttrsatcka gagctgaaat ctggatagta cctgcaggaa caatattcct    1116
gtagccatgg aagagggcca aggctcagtc actccttgga tggcctccta aatctccccg     1176
tggcaacagg tccaggagag gcccatggag cagtctcttc catggagtaa gaaggaaggg     1236
agcatgtacg cttggtccaa gattggctag ttccttgata gcatcttact ctcaccttct     1296
ttgtgtctgt gatgaaagga acagtctgtg caatgggttt tacttaaact tcactgttca     1356
acctatgagc aaatctgtat gtgtgagtat aagttgagca tagcatactt ccagaggtgg     1416
tcttatggag atggcaagaa aggaggaaat gatttcttca gatctcaaag gagtctgaaa     1476
tatcatattt ctgtgtgtgt cdctctcagc ccctgcccad gctagaggga wacagctact     1536
gataatcgaa aactgctgtt tgtgggcarg aacccctggc tgtgcaaata atggggctga     1596
ngccctgtgt gatattgaaa aaaaaaa                                          1623
```

<210> 319
<211> 526
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 99..371

<221> sig_peptide
<222> 99..290
<223> Von Heijne matrix
      score 3.79999995231628
      seq LFIVVCVICVTLN/FP

<221> polyA_signal
<222> 491..496

<221> polyA_site
<222> 513..524

<400> 319
```
attggattag tagaattgct tttgtcattc cattgttttc atatatttgt ttgggacatt       60
ttacttttttt ctgttaacgc ttaccctagr aattagaa atg aca cca cgt att ctt      116
                                          Met Thr Pro Arg Ile Leu
                                                                -60
agc gaa gtc cag ttt tca gca ttt tgt cct tat tgg aca ata gca agg        164
Ser Glu Val Gln Phe Ser Ala Phe Cys Pro Tyr Trp Thr Ile Ala Arg
        -55                 -50                 -45
ata tta gaa cgt gtt ggt tcc gcg tgc ttc cgt ctt gag tta tgt gct        212
Ile Leu Glu Arg Val Gly Ser Ala Cys Phe Arg Leu Glu Leu Cys Ala
        -40                 -35                 -30
gct att gtc gga tat ttt gtc tta gat gta cgt act ttc ctg ttc att        260
Ala Ile Val Gly Tyr Phe Val Leu Asp Val Arg Thr Phe Leu Phe Ile
        -25                 -20                 -15
gtg gta tgt gta att tgc gtt act ttg aat ttt cca cgt ttt tac ttt        308
Val Val Cys Val Ile Cys Val Thr Leu Asn Phe Pro Arg Phe Tyr Phe
-10                      -5                  1                  5
ctt tgt ctc tca tca ctt acc gct ttt ggg acc ccc ccc atc ggg gtt        356
Leu Cys Leu Ser Ser Leu Thr Ala Phe Gly Thr Pro Pro Ile Gly Val
                10                  15                  20
cac att ccc tct ccc tararcacac tcccttggat ttcctcradt ggggtctgct        411
His Ile Pro Ser Pro
                25
gcggtgaagc tttcccattt tatgtgcaga ttattttcag agggtatata gaattcaggc      471
agctgtttcg ttgtagcaca ttaaaaatat tttcccactt caaaaaaaaa aaacc          526
```

<210> 320
<211> 989
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 44..814

<221> sig_peptide
<222> 44..112
<223> Von Heijne matrix
      score 8.30000019073486
      seq VRLLLXLLLLLIA/LE

<221> polyA_site
<222> 978..989

<400> 320
```
aaatgtgtac acgcccagct tcctgcctgt tactctccac agt atg cga aga ata          55
                                              Met Arg Arg Ile
                                                          -20
tcc ctg act tct agc cct gtg cgc ctt ctt ttg tdt ctg ctg ttg cta        103
Ser Leu Thr Ser Ser Pro Val Arg Leu Leu Leu Xaa Leu Leu Leu Leu
              -15                       -10                   -5
cta ata gcc ttg gag atc atg gtt ggt ggt cac tct ctt tgc ttc aac        151
Leu Ile Ala Leu Glu Ile Met Val Gly Gly His Ser Leu Cys Phe Asn
             1               5                  10
ttc act ata aaa tca ttg tcc aga cct gga cag ccc tgg tgt gaa gcg        199
Phe Thr Ile Lys Ser Leu Ser Arg Pro Gly Gln Pro Trp Cys Glu Ala
     15              20                  25
cat gtc ttc ttg aat aaa aat ctt ttc ctt cag tac aac agt gac aac        247
His Val Phe Leu Asn Lys Asn Leu Phe Leu Gln Tyr Asn Ser Asp Asn
30                  35                  40                  45
aac atg gtc aaa cct ctg ggc ctc ctg ggg aag aag gta tat gcc acc        295
Asn Met Val Lys Pro Leu Gly Leu Leu Gly Lys Lys Val Tyr Ala Thr
                 50                  55                  60
agc act tgg gga gaa ttg acc caa acg ctg gga gaa gtg ggg cga gac        343
Ser Thr Trp Gly Glu Leu Thr Gln Thr Leu Gly Glu Val Gly Arg Asp
                 65                  70                  75
ctc agg atg ctc ctt tgt gac atc aaa ccc car ata aag acc agt gat        391
Leu Arg Met Leu Leu Cys Asp Ile Lys Pro Gln Ile Lys Thr Ser Asp
             80                  85                  90
cct tcc act ctg caa gtc kar atk ttt tgt caa cgt gaa gca gaa cgg        439
Pro Ser Thr Leu Gln Val Xaa Xaa Phe Cys Gln Arg Glu Ala Glu Arg
             95                  100                 105
tgc act ggt gca tcc tgg cag ttc gcc acc aat gga gag aaa tcc ctc        487
Cys Thr Gly Ala Ser Trp Gln Phe Ala Thr Asn Gly Glu Lys Ser Leu
110                 115                 120                 125
ctc ttt gac gca atg aac atg acc tgg aca gta att aat cat gaa gcc        535
Leu Phe Asp Ala Met Asn Met Thr Trp Thr Val Ile Asn His Glu Ala
                 130                 135                 140
agt wag atc aag gag aca tgg aag aaa gac aga ngg ctg gaa aak tat        583
Ser Xaa Ile Lys Glu Thr Trp Lys Lys Asp Arg Xaa Leu Glu Xaa Tyr
                 145                 150                 155
ttc agg aag ctc tca aar gga gac tgc gat cac tgg ctc agg gaa ttc        631
Phe Arg Lys Leu Ser Lys Gly Asp Cys Asp His Trp Leu Arg Glu Phe
             160                 165                 170
tta ggg cac tgg gaa gca atg cca raa ccg ama gtg tcm cca rta aat        679
```

```
Leu Gly His Trp Glu Ala Met Pro Xaa Pro Xaa Val Ser Pro Xaa Asn
    175                 180                 185
gct tca raw atc cac tgg tct tct tct art cta cca raw ara tgg atc    727
Ala Ser Xaa Ile His Trp Ser Ser Ser Xaa Leu Pro Xaa Xaa Trp Ile
190                 195                 200                 205
atc ctg ggg gca ttc atc ctg tta vtt tta atg gga att gtt ctc atc    775
Ile Leu Gly Ala Phe Ile Leu Leu Xaa Leu Met Gly Ile Val Leu Ile
                210                 215                 220
tgt gtc tgg tgg caa aat ggc ara ara tcc acc tad arg tgataccacg     824
Cys Val Trp Trp Gln Asn Gly Xaa Xaa Ser Thr Xaa Xaa
                225                 230
gcggcgcaaa attgttcacc tgtggtcctc gatcgctgac agccttggct cccactgctg   884
tgtgttccct gagtcaagtg gaggcggagc ctgcaatgag cggaratcgc gcctctgcat   944
tccagtcttg caacagarc aagactccgt ctcaaaaaaa aaaaa                    989
```

```
<210> 321
<211> 1017
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 3..581

<221> sig_peptide
<222> 3..182
<223> Von Heijne matrix
      score 6.69999980926514
      seq LWPFLTWINPALS/IC

<221> polyA_site
<222> 1006..1016

<400> 321
ac atg tgc cct agt ctg gaa gag gct ccc agt gtc aag ggg act ctg      47
   Met Cys Pro Ser Leu Glu Glu Ala Pro Ser Val Lys Gly Thr Leu
   -60                 -55                 -50
ccc tgc tca gga caa cag cag cct ttc ccg ttt gga gcc tca aac atc     95
Pro Cys Ser Gly Gln Gln Gln Pro Phe Pro Phe Gly Ala Ser Asn Ile
-45                 -40                 -35                 -30
cca cta ctc ctg ggc agg agc aga aag gtg gct cga ggt gca ccg gtc    143
Pro Leu Leu Leu Gly Arg Ser Arg Lys Val Ala Arg Gly Ala Pro Val
                -25                 -20                 -15
ctg tgg cca ttt ctc act tgg ata aac cct gca ctg tcc atc tgt gac    191
Leu Trp Pro Phe Leu Thr Trp Ile Asn Pro Ala Leu Ser Ile Cys Asp
                -10                 -5                  1
ccc tta gga tcc tgc gga tgg cyw tgc cac acg gcc car gtc cct gcg    239
Pro Leu Gly Ser Cys Gly Trp Xaa Cys His Thr Ala Gln Val Pro Ala
    5                   10                  15
ccc ctg car ttg cct act gcc tgt cct ccc ctc cca cat ggc acc cgg    287
Pro Leu Gln Leu Pro Thr Ala Cys Pro Pro Leu Pro His Gly Thr Arg
20                  25                  30                  35
gct gta ggc ccc acg cca ggc ctc ctc cct gag gct gca gcc cca sgc    335
Ala Val Gly Pro Thr Pro Gly Leu Leu Pro Glu Ala Ala Ala Pro Xaa
                40                  45                  50
acg tgk ggg gca ctg tcc tca cgc agc agg cac tgg tca tgt tcc att    383
Thr Xaa Gly Ala Leu Ser Ser Arg Ser Arg His Trp Ser Cys Ser Ile
                55                  60                  65
gtc arc tgc ctc cac ctg cac ara ctc ctg tct gtg gag acc aga arc    431
Val Xaa Cys Leu His Leu His Xaa Leu Leu Ser Val Glu Thr Arg Xaa
```

```
          70                    75                    80
ttc cas aaa cat ctg ttg gtg ctg ctg gtg gct gtg gcc cat agt gtt      479
Phe Xaa Lys His Leu Leu Val Leu Leu Val Ala Val Ala His Ser Val
       85                    90                    95
ctg gaa cca cct gcc ctg gtc cca aat gtg cag tgt gag atg tgc aca      527
Leu Glu Pro Pro Ala Leu Val Pro Asn Val Gln Cys Glu Met Cys Thr
100                   105                   110                   115
cac tca ggg ccc cgt gac ctg gaa gcc gca gtc gtg tcc cca gca cct      575
His Ser Gly Pro Arg Asp Leu Glu Ala Ala Val Val Ser Pro Ala Pro
                    120                   125                   130
tgg gaa tgagcctgtc ctctgtgtga aggagggggt ggttctcaaa ccactgactc      631
Trp Glu
ttggtgctca ggaggggcct gctgctgtcc tggcatggg gtggtcattg ttcaagactg      691
aggcagactc agtctttgaa agggtgcaga ggccaggcgc ggtggctcac gcctgtaatt      751
ccagcacttt gggaggccaa ggtggacaga tcatgaggtc aggagttcga gaccagcctg      811
gccaatacgg tgaaaccgca tctctactaa rraatawcaw aaattagtcg ggcatgggtg      871
atgtgtgctt gtagtcccag ctactcatga ggyctgaggc agaagaatca cctgaatctg      931
ggaggcagag gttgcagtga accaagatcg cacgactgta caccagcctg ggcgacagag      991
tgagactccg tctcaaaaaa aaaaam                                          1017
```

```
<210> 322
<211> 529
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 107..427

<221> sig_peptide
<222> 107..190
<223> Von Heijne matrix
      score 3.79999995231628
      seq RFLSLSAADGSDG/SH

<221> polyA_signal
<222> 499..504

<221> polyA_site
<222> 516..529

<400> 322
aaagtcagcg ctggagtcgg ctaggcggct ggaaacggcg gctgccgccg gtgactcagg      60
gaggcgggag gccgmsggmg gagctcttcc tgcaggcgtg garacc atg gtg ctc      115
                                                    Met Val Leu
acg ctc gga gaa agt tgg ccg gta ttg gtg ggg agg agg ttt ctc agt      163
Thr Leu Gly Glu Ser Trp Pro Val Leu Val Gly Arg Arg Phe Leu Ser
-25                   -20                   -15                   -10
ctg tcc gca gcc gac ggc agc gat ggc agc cac gac agc tgg gac gtg      211
Leu Ser Ala Ala Asp Gly Ser Asp Gly Ser His Asp Ser Trp Asp Val
                    -5                    1                     5
gag cgc gtc gcc gag tgg ccc tgg ctc tcc ggg acc att cga gct gtt      259
Glu Arg Val Ala Glu Trp Pro Trp Leu Ser Gly Thr Ile Arg Ala Val
                    10                    15                    20
tcc cac acc gac gtt acc aag aag gat ctg aag gtg tgt gtg gaa ttt      307
Ser His Thr Asp Val Thr Lys Lys Asp Leu Lys Val Cys Val Glu Phe
      25                    30                    35
gak ggg gaa tct tgg agg aaa aga aga tgg ata gaa gtc tac agc ctt      355
Xaa Gly Glu Ser Trp Arg Lys Arg Arg Trp Ile Glu Val Tyr Ser Leu
40                    45                    50                    55
```

```
cta agg aaa gca ttt tta gta aaa cat aat ttg gtt tta gct gaa cga    403
Leu Arg Lys Ala Phe Leu Val Lys His Asn Leu Val Leu Ala Glu Arg
            60                  65                  70
aag tca cct gaa att tct tgg ggt taaccatctt tagttaaatg gaattttaat    457
Lys Ser Pro Glu Ile Ser Trp Gly
            75
ttaaatgacg ctttgctaat tttaagtgtt aagcattttg cattaaaata ttcatataat    517
aaaaaaaaaa aa                                                        529
```

```
<210> 323
<211> 1046
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 45..407

<221> sig_peptide
<222> 45..83
<223> Von Heijne matrix
      score 5.69999980926514
      seq MLVLRSALTRALA/SR

<221> polyA_signal
<222> 1008..1013

<221> polyA_site
<222> 1032..1042

<400> 323
aaaaggacac ggctggctgc ttttctcagc gccgaagccg cgcc atg ctc gtc ctc     56
                                              Met Leu Val Leu
                                                            -10
aga agc gcc ctg act cgg gcg ctg gcc tca cgg acg ctg gcg cct cag    104
Arg Ser Ala Leu Thr Arg Ala Leu Ala Ser Arg Thr Leu Ala Pro Gln
            -5                  1                   5
atg tgc tca tct ttt gct acg gga ccc aga caa tac gat gga ata ttc    152
Met Cys Ser Ser Phe Ala Thr Gly Pro Arg Gln Tyr Asp Gly Ile Phe
            10                  15                  20
tat gaa ttt cgt tct tat tac ctt aag ccc tca aag atg aat gag ttc    200
Tyr Glu Phe Arg Ser Tyr Tyr Leu Lys Pro Ser Lys Met Asn Glu Phe
    25                  30                  35
ctg gaa aat ttt gag aaa aac gct caa ctt cgg aca gct cac tct gaa    248
Leu Glu Asn Phe Glu Lys Asn Ala Gln Leu Arg Thr Ala His Ser Glu
40                  45                  50                  55
ttg gtt gga tac tgg agt gta kaa ttt gga ggc aga atg awt aca gtg    296
Leu Val Gly Tyr Trp Ser Val Xaa Phe Gly Gly Arg Met Xaa Thr Val
            60                  65                  70
ttt cat att tgg aag tat gat aat ttt gct cat cga act gaa ttt cag    344
Phe His Ile Trp Lys Tyr Asp Asn Phe Ala His Arg Thr Glu Phe Gln
            75                  80                  85
aaa gcc ttg gcc aaa gat aag gaa tgg caa gaa caa ttc ctc att cca    392
Lys Ala Leu Ala Lys Asp Lys Glu Trp Gln Glu Gln Phe Leu Ile Pro
            90                  95                  100
aat ttg gct ctc aat tgataaacaa gatagtgaga ttacttatct ggtaccatgg    447
Asn Leu Ala Leu Asn
            105
tgcaaattag aaaaacctcc aaaagaagga gtctatgaac tggccacttt tcagatgaaa    507
cctggtgggc cagctctgtg gggtgatgca tttaaaaggg cagttcatgc tcatgtcaat    567
```

```
ctaggctaca caaaactagt tggagtgttc cacacagagt acggagcact caacagagtt    627
catgttcttt ggtggaatga gagtgcagat agtcgtgcag ctgggagaca taagtcccat    687
gaggatccca gagttgtggc agctgttcgg gaaagtgtca actacctagt atctcagcag    747
aatatgcttc tgattcctac atcgttttca ccactgaaat agttttctac tgaaatacaa    807
aacattcat taactgctat aggatctgtc tgctaatggt gcttaaattc tcccaagagg    867
ttctcacttt tatttgaagg aggtggtaag ttaatttgct atgtttcttg cattatgaag    927
gctacatctg tgctttgtaa gtaccacttc aaaaaatakt tctgtttact ttctgcatgg    987
tatttcagtg tctgtcatac attaaaaata cttgtcactg tttyaaaaaa aaaaammcc    1046
```

```
<210> 324
<211> 880
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 201..332

<221> sig_peptide
<222> 201..251
<223> Von Heijne matrix
      score 7.80000019073486
      seq VLWLISFFTFTDG/HG

<221> polyA_site
<222> 869..880

<400> 324
aattgctgat ggatcagtga gcctgtgttc atgccagtga gctgctgtgg ctcagatact     60
gatactttct ttccaaacag cataagaagt gattgancca caagtatact gaaggmargg    120
yhcccwsvar tyctggwgtg amgagataaa tcaccagtca cagactatgc acccgactgc    180
tgctgttcag tccagggaaa atg aaa gtt gga gtg ctg tgg ctc att tct ttc    233
                        Met Lys Val Gly Val Leu Trp Leu Ile Ser Phe
                        -15                       -10
ttc acc ttc act gac ggc cac ggt ggc ttc ctg ggg gtg agt tgg tgc    281
Phe Thr Phe Thr Asp Gly His Gly Gly Phe Leu Gly Val Ser Trp Cys
 -5              1             5                    10
tat gtc tca tat ctc ttc tca act aac tct cct ctc tcg ttc cgg cgc    329
Tyr Val Ser Tyr Leu Phe Ser Thr Asn Ser Pro Leu Ser Phe Arg Arg
           15                20  .                 25
att tagaacccct cactctctag gggactgcaa ctgcataatt taatgtactt    382
Ile
gagatcagaa gtcctgagtt ctcgtttcaa cattaccaac attcactgtg tggccttgga    442
taagtragtc atttcatctc ttcggagctt agatgatcma actgcaarag gaggatcttt    502
gattamacta tcttagagat ctttttccagt tcaacacatg ctgtactatg gcttctcgga    562
tgcagaaaaa tcacatggat ggacattagc aatccttara cactgtcttt cctgtctaca    622
ctcgcttgag tgatgckttc atctaggatc atggtttttaa tattctctac atgctgatga    682
ctcccagctg tatagctcca tctcagaacc tctccctgt ccacactcac atatccatta    742
cctacgtgtt atttccagct gggaaatcca gcggaacctc ggnaacttca tttgnttcaa    802
aatcgnaacc caatccttct tgcctatctc agcaagtggt atcactatct ttccagctac    862
ttaggcaaaa aaaaaaaa                                                   880
```

```
<210> 325
<211> 1217
<212> DNA
<213> Homo sapiens

<220>
```

<221> CDS
<222> 217..543

<221> sig_peptide
<222> 217..255
<223> Von Heijne matrix
      score 6.40000009536743
      seq MCLLTALVTQVIS/LR

<221> polyA_site
<222> 1206..1217

<400> 325
```
aatgccagtg tcagcttctc tccgaaaact gggtaatacg aaatggtctt tattggttgt      60
gaacactcga gctgagaaac attttaggat ctttgtgtct tttgtgatga ttttgtttct     120
graagrwgga aasctgtcta aaaatattca agtgtgcaac caaggattta gatgaagcca     180
gcaaacaaag gaatcatgta atcaggacct gagcga atg tgc tta ctc acg gcg        234
                                          Met Cys Leu Leu Thr Ala
                                                            -10

tta gtt aca cag gtg att tcc tta aga aaa aat gca gag aga act tgt        282
Leu Val Thr Gln Val Ile Ser Leu Arg Lys Asn Ala Glu Arg Thr Cys
        -5                  1                   5

tta tgc aag agg aga tgg ccc tgg ngc ccc tcg ccc cgg atc tac tgc        330
Leu Cys Lys Arg Arg Trp Pro Trp Xaa Pro Ser Pro Arg Ile Tyr Cys
10                  15                  20                  25

tca tcc acc cca tgc gat tcc aaa ttc ccc acc gtc tac tcc agt gcc        378
Ser Ser Thr Pro Cys Asp Ser Lys Phe Pro Thr Val Tyr Ser Ser Ala
                30                  35                  40

cca ttc cat gcc ccc ctc ccc gtc cag aat tcc tta tgg ggg cac ccg        426
Pro Phe His Ala Pro Leu Pro Val Gln Asn Ser Leu Trp Gly His Pro
            45                  50                  55

ctc cat ggt tgt tcc tgg caa tgc cac cat ccc cag gga car aat ctc        474
Leu His Gly Cys Ser Trp Gln Cys His His Pro Gln Gly Gln Asn Leu
            60                  65                  70

cag cct gcc agt ctc cad acc cat ctc tcc aag ccc aag cgc cat ttt        522
Gln Pro Ala Ser Leu Xaa Thr His Leu Ser Lys Pro Lys Arg His Phe
        75                  80                  85

ara aar aar rra tgt caa gcc tgatgaarac atgagtggca aaaacattgc           573
Xaa Lys Lys Xaa Cys Gln Ala
90                  95

aatgtacara aatgagggtt tctatgctga tccttacctt tatcacgagg gacggatgag      633
catasccTca tcccatggtg gacacccact ggatgtcccc gaccacatca ttgcatatca      693
ccgcaccgcc atccggtcag cgagtgctta ttgtaacccc tcaatgcaag cggaaatgca      753
tatggaacaa tcactgtaca gacagaaatc aaggaaatat ccggatagcc atttgcctac      813
actgggctcc aaaacacccc ctgcctctcc tcacagaktc agtgacctga ggatgataga      873
catgcacgct cactataatg cccacggccc ccctcacacc atgcagccag accgggcctc      933
tccgagccgc caggccttta aaaaggagcc aggcaccttg gtgtatatag aaaagccacg      993
gagcgctgca ggattatcca gccttgtaga cctcggccct cctctaatgg agaagcaagt     1053
ttttgcctac agcacggcga caatacccaa agacagagag accagagaga ggatgcaagc     1113
catggagaaa cagattgcca gtttaactgg ccttgttcag tctgcgcttt ttaaagggcc     1173
cattacaagt tatagcaaar atgcgtctag ctaaaaaaaa aaaa                      1217
```

<210> 326
<211> 959
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 18..446

```
<221> sig_peptide
<222> 18..140
<223> Von Heijne matrix
      score 4.09999990463257
      seq GILILWIIRLLFS/KT

<221> polyA_signal
<222> 930..935

<221> polyA_site
<222> 948..959

<400> 326
aaaggaagcg gctaact atg gcg acc gcc acg gag cag tgg gtt ctg gtg          50
                    Met Ala Thr Ala Thr Glu Gln Trp Val Leu Val
                    -40                  -35
gag atg gta cag gcg ctt tac gag gct cct gct tac cat ctt att ttg         98
Glu Met Val Gln Ala Leu Tyr Glu Ala Pro Ala Tyr His Leu Ile Leu
-30                  -25                  -20                  -15
gaa ggg att ctg atc ctc tgg ata atc aga ctt ctt ttc tct aag act        146
Glu Gly Ile Leu Ile Leu Trp Ile Ile Arg Leu Leu Phe Ser Lys Thr
                -10                  -5                        1
tac aaa tta caa gaa cga tct gat ctt aca gtc aag gaa aaa gaa gaa        194
Tyr Lys Leu Gln Glu Arg Ser Asp Leu Thr Val Lys Glu Lys Glu Glu
         5                   10                  15
ctg att gaa gag tgg caa cca gaa cct ctt gtt cct cct gtc cca aaa        242
Leu Ile Glu Glu Trp Gln Pro Glu Pro Leu Val Pro Pro Val Pro Lys
     20                  25                  30
gac cat cct gct ctc aac tac aac atc gtt tca ggc cct cca agc cac        290
Asp His Pro Ala Leu Asn Tyr Asn Ile Val Ser Gly Pro Pro Ser His
35                  40                   45                   50
aaa act gtg gtg aat gga aaa gaa tgt ata aac ttc gcc tca ttt aat        338
Lys Thr Val Val Asn Gly Lys Glu Cys Ile Asn Phe Ala Ser Phe Asn
                55                   60                  65
ttt ctt gga ttg ttg gat aac cct agg gtt aag gca gca gct tta gca        386
Phe Leu Gly Leu Leu Asp Asn Pro Arg Val Lys Ala Ala Ala Leu Ala
             70                  75                  80
tct cta aag aag tat ggc gtg ggg act tgt gga ccc tgt gga ttt tat        434
Ser Leu Lys Lys Tyr Gly Val Gly Thr Cys Gly Pro Cys Gly Phe Tyr
         85                  90                  95
ggc aca ttt gaa tgaaatgaa ggatcattga tttccttgtg tatggataat           486
Gly Thr Phe Glu
         100

ccgggaacag gccaactaaa tatttgatga atgtatgatt tcaaatacag tgaattccct      546
gggagtcatc aaaraagacg gcattttatg gttgttttta ttaagtgtat attctttgct      606
cctgaaaatg ttattaaata attgtttagg ccgggcatgg tggctcatgc ctgtaatccc      666
agcactttca aaggctgagg caggcagatc acctgaggtc aggagttcaa aaccagcctg      726
gccaacatgc tgaaacctcg tctctactaa aaatacaaaa attagctggg cgtggtggtg      786
grtgcctgtg gtcccagctr cgtgggaggc tgaggtggga gaattgcttc aacctgggag      846
gcggaggttg cagtgagccg agatcatgcc actgcactcc agcctgggca acagagcaag      906
actgtctcaa aaataaataa ataaataaaa ttgtttaaat gaaaaaaaaa aaa             959



<210> 327
<211> 921
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
```

<222> 29..724

<221> sig_peptide
<222> 29..118
<223> Von Heijne matrix
      score 3.90000009536743
      seq VAHALSLPAESYG/NX

<221> polyA_signal
<222> 886..891

<221> polyA_site
<222> 910..920

<400> 327
```
aaggagccac gctttcgggg gttgcaag atg gcg gcc acc agt gga act gat        52
                                 Met Ala Ala Thr Ser Gly Thr Asp
                                 -30               -25

gag ccg gtt tcc ggg gag ttg gtg tct gtg gca cat gcg ctt tct ctc     100
Glu Pro Val Ser Gly Glu Leu Val Ser Val Ala His Ala Leu Ser Leu
    -20               -15               -10

cca gca gag tcg tat ggy aac grt yct gac att gag atg gct tgg gcc     148
Pro Ala Glu Ser Tyr Gly Asn Xaa Xaa Asp Ile Glu Met Ala Trp Ala
    -5                1               5               10

atg aga gca atg cag cat gct gaa gtc tat tac aag ctg att tca tca     196
Met Arg Ala Met Gln His Ala Glu Val Tyr Tyr Lys Leu Ile Ser Ser
            15               20               25

gtt gac cca cag ttc ctg aaa ctc acc aaa gta gat gac caa att tac     244
Val Asp Pro Gln Phe Leu Lys Leu Thr Lys Val Asp Asp Gln Ile Tyr
            30               35               40

tct gag ttc cgg aaa aat ttt gag acc ctt agg ata gat gtg ttg grc     292
Ser Glu Phe Arg Lys Asn Phe Glu Thr Leu Arg Ile Asp Val Leu Xaa
            45               50               55

cca gaa gan ctc aag tca gaa tca gcn aaa gag ccc cca gga tac aat     340
Pro Glu Xaa Leu Lys Ser Glu Ser Ala Lys Glu Pro Pro Gly Tyr Asn
    60               65               70

tct ttg cca ttg aaa ttg ctc gga acc ggg aag gct ata aca aag ctg     388
Ser Leu Pro Leu Lys Leu Leu Gly Thr Gly Lys Ala Ile Thr Lys Leu
75               80               85               90

ttt ata tca gtg ttc agg aca aag aag gag aga aag gag tca aca atg     436
Phe Ile Ser Val Phe Arg Thr Lys Lys Glu Arg Lys Glu Ser Thr Met
            95               100               105

gag gag aaa aaa gag ctg aca gtg gag aag aag aga aca cca aga atg     484
Glu Glu Lys Lys Glu Leu Thr Val Glu Lys Lys Arg Thr Pro Arg Met
            110               115               120

gag gag aga aag gag ctg ata gtg gag aag aaa aag agg aag gaa tca     532
Glu Glu Arg Lys Glu Leu Ile Val Glu Lys Lys Lys Arg Lys Glu Ser
            125               130               135

aca gag aag aca aaa ctg aca aag gag gag aaa aag gga aag aag ctg     580
Thr Glu Lys Thr Lys Leu Thr Lys Glu Glu Lys Lys Gly Lys Lys Leu
            140               145               150

aca aag aaa tca aca aaa gtg gtg aaa aag cta tgt aag gta tac agg     628
Thr Lys Lys Ser Thr Lys Val Val Lys Lys Leu Cys Lys Val Tyr Arg
155               160               165               170

gaa cag cac tct aga agc tat gac tca att gag act aca agt acc acg     676
Glu Gln His Ser Arg Ser Tyr Asp Ser Ile Glu Thr Thr Ser Thr Thr
            175               180               185

gtg cta ctt gca cag acc cct ttg gtt aaa tgt aaa ttc ttg tac aat     724
Val Leu Leu Ala Gln Thr Pro Leu Val Lys Cys Lys Phe Leu Tyr Asn
            190               195               200

tgaaggatac gcagaaggac atcttctag tctaacagtc aggagctgct ctggtcattc     784
ccttgtatga actggtctaa agactgttag tggggtgtta gttgattttt cctggtatac     844
```

```
tgtttcttgg ctgacactac tggtcaagta agaaatttgt aaataaattt cttttggttc      904
ttattaamaa aaaaaas                                                      921
```

<210> 328
<211> 1344
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 404..586

<221> sig_peptide
<222> 404..466
<223> Von Heijne matrix
     score 4.09999990463257
     seq SLMFFSMMATCTS/NV

<221> polyA_signal
<222> 1304..1309

<221> polyA_site
<222> 1334..1344

<400> 328
```
ataatttaat gcaaaatatc ctttttatgaa tttcatgtta atattgtgaa atattaaaat      60
aattccacaa tagttgagaa aaatgagcat ttttttccat ttttaaaaaa tgcatagaaa      120
agacaatttt aaaatcctgg gamccawatt tatttagaag tagctgttag taaaacatta      180
gaaaaggagt caggccatba ggttatttat nbnaatctct aagcaattag gntgaagtta      240
ttaagtcaag cctagaaaag ctgcctcctt gtaaggcttt catgacaatg tatagtaatc      300
brcagtgtcc aattcttcgc actcctcagg aatatcacta cctcaggtta cggtacacag      360
gctataattg atgatgatgt tcagataact gaagacacaa taa atg aca ttc aga      415
                                                 Met Thr Phe Arg
                                                 -20
cat cag gac aat tcc ctc atg ttc ttt tct atg atg gcc acc tgt acc      463
His Gln Asp Asn Ser Leu Met Phe Phe Ser Met Met Ala Thr Cys Thr
    -15                -10                -5
agc aac gtg ggt ttc acc cac aca acg atg aac tgt tct ctt act tct      511
Ser Asn Val Gly Phe Thr His Thr Thr Met Asn Cys Ser Leu Thr Ser
    1           5               10              15
cca gtt gat ttt aaa gac ttg tta aga gtc tta cta ata aaa ttt ggg      559
Pro Val Asp Phe Lys Asp Leu Leu Arg Val Leu Leu Ile Lys Phe Gly
            20              25              30
tat gat aga aaa tcc aca atc aaa tct tgaaccaaat aacatattaa            606
Tyr Asp Arg Lys Ser Thr Ile Lys Ser
            35              40
attactaata tttaagtgat ggaagacaca caaaaaactt aaaagcacga acaacctaac      666
ttgaaaaara attttaaaat atgattaacc tgaaraaaar araatcctaa ragccaaagc      726
tccttttat ttagcttgga attttcctat tggttcctaa caaactgtcc caatgtcata      786
taaggaaaca tgatctatta cattccttta taacaacgtg gararactat aaacctatgt      846
aagtagtaaa actatatcag adactcagga ractgactww aaggcctgga tctgcagtgt      906
attatctgta taaaaattgg caggggggaag ctaaaaggaa aggagattgg agatctcaat      966
tctatcatgg tgtatttcat acgcaaatca ragcatgcat tgttttttgt ttttggaaar      1026
avaarggaag tgtgttctgc cccatgtttc cttccgtgtt tatagttcaa actctatata      1086
tacttcaggt atttttttgtt tagcccttca ttataaatgg gcaggaaatt gtttatcaac      1146
ctagccagtt tattactagt gaccttgact tcagtatctt gagcattctt ttatattttt      1206
ctttttattat cctgagtctg taactaaaca attttgtctt caaattttta tccaatatcc      1266
attgcaccac accaaatcaa gcttcttgat tttcaaaaat aaaaaggggg aaatacttac      1326
aacttgtaaa aaaaaaaa                                                     1344
```

```
<210> 329
<211> 585
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 331..432

<221> sig_peptide
<222> 331..387
<223> Von Heijne matrix
      score 7
      seq AGLSSCLLPLCWL/ER

<221> polyA_signal
<222> 548..553

<221> polyA_site
<222> 573..585

<400> 329
aagcctaggt gtggcgcccc gaccggactt tcacttctgg ccagcccttt ccccacctgg    60
gcgcgggass ggtgccagtc tttaaacaac ctctcgatgg tcccacgaa gatgtttcca   120
gacccttgga atgccaagtt caagtttagc tatgtctcgc ggagaggccg gtggaagaag   180
caacgagaat gaagcacccc agttctctgc tgagcacatg ggcatctgca ataaagattt   240
aatttcccag cttctcctga agctcggtat ggccacaaca ctaaattctg cccgaggaga   300
ttgagcaaaa tagtatggga cttccaagaa atg ttt tta aag tca ggg gca ggc   354
                                  Met Phe Leu Lys Ser Gly Ala Gly
                                                                -15
ctt tct tca tgc ctt ctt cct ctt tgc tgg ctg gaa cgc aaa gac cat   402
Leu Ser Ser Cys Leu Leu Pro Leu Cys Trp Leu Glu Arg Lys Asp His
    -10               -5                  1                   5
ggc agg agg cca agc asc cat cct gga agg tgaaagcctc atactaagga       452
Gly Arg Arg Pro Ser Xaa His Pro Gly Arg
                10          15
cgtcaracag cgaaataara rcctgggtcc ttgaccctgt aaasatctcc ctccccatcc   512
tggtctgtct gccttgactc ctttcatatg aaaaaaataa acttttaact tgcgtwaacc   572
aaaaaaaaaa aaa                                                      585
```

```
<210> 330
<211> 914
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 59..703

<221> sig_peptide
<222> 59..220
<223> Von Heijne matrix
      score 5.09999990463257
      seq FLLSQMSQHQVHA/VQ

<221> polyA_signal
<222> 886..891
```

<221> polyA_site
<222> 903..914

<400> 330
acaaatatca atgatgttta tgaatctagt gtgaaagtkt taatcacatc acaaggct   58
atg aac rra tat gca agt cca ttc aac tgw caa ttg ard tat ttg gak   106
Met Asn Xaa Tyr Ala Ser Pro Phe Asn Xaa Gln Leu Xaa Tyr Leu Xaa
                -50   -45   -40
ttg agc agr ttc gag tgt gtr cat aga gat gga aga gta att aca ctg   154
Leu Ser Arg Phe Glu Cys Val His Arg Asp Gly Arg Val Ile Thr Leu
        -35   -30   -25
tct tat cag gag cag gag cta cag gat ttt ctt ctg tct cag atg tca   202
Ser Tyr Gln Glu Gln Glu Leu Gln Asp Phe Leu Leu Ser Gln Met Ser
        -20   -15   -10
cag cac cag gta cat gca gtt cag caa ctc gcc aag gtt atg ggc tgg   250
Gln His Gln Val His Ala Val Gln Gln Leu Ala Lys Val Met Gly Trp
    -5   1   5   10
caa gta ctg agc ttc agt aat cat gtg gga ctt gga cct ata gag agc   298
Gln Val Leu Ser Phe Ser Asn His Val Gly Leu Gly Pro Ile Glu Ser
                15   20   25
abt ggt aat gca tct gcc atc acg gtg gcc ccc caa gtg gtg act atg   346
Xaa Gly Asn Ala Ser Ala Ile Thr Val Ala Pro Gln Val Val Thr Met
            30   35   40
cta ttt cag ttc gta atg gac ctg aaa gtg gca gca aga tta tgg ttc   394
Leu Phe Gln Phe Val Met Asp Leu Lys Val Ala Ala Arg Leu Trp Phe
        45   50   55
agt ttc ctc gta acc aat gta aar acc ttc caa aaa gtg atg ttt tac   442
Ser Phe Leu Val Thr Asn Val Lys Thr Phe Gln Lys Val Met Phe Tyr
    60   65   70
aar ata aca aat gga gtc atc ttc gtg ggc cat tca aar aag ttc agt   490
Lys Ile Thr Asn Gly Val Ile Phe Val Gly His Ser Lys Lys Phe Ser
75   80   85   90
gga ata aaa tgg aag gtc kaa att ttg ttt ata aaa tgg arm tgc tta   538
Gly Ile Lys Trp Lys Val Xaa Ile Leu Phe Ile Lys Trp Xaa Cys Leu
                95   100   105
tgt ctg cac tta gcc ctt gtc tac tat gat ttt ttc car atg ttt cct   586
Cys Leu His Leu Ala Leu Val Tyr Tyr Asp Phe Phe Gln Met Phe Pro
            110   115   120
aaa raa gtt tcc ara aac ttt gac ttg aaa tgt ttg car atc aac tat   634
Lys Xaa Val Ser Xaa Asn Phe Asp Leu Lys Cys Leu Gln Ile Asn Tyr
        125   130   135
aag cac aaa gaa gar ata act tcc aaa aga gtg ctg ttt tta aaa ata   682
Lys His Lys Glu Glu Ile Thr Ser Lys Arg Val Leu Phe Leu Lys Ile
    140   145   150
ata att agg aaa tgt ttt att tagcactttc aaactttca ctttataaat   733
Ile Ile Arg Lys Cys Phe Ile
155   160
gacaagtgct ttgaaatgca gaagtttatg tacagttgta tatacagtat gacaagatgt   793
aaaataatat gtttttcatg cagtttaaaa tattactaac ttaagggttt ctatgtgctt   853
tttaaaatat tccttctttg atgttgacat caaataaagt atgtggttta aaaaaaaaaa   913
a   914

<210> 331
<211> 1161
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 672..752

```
<221> sig_peptide
<222> 672..722
<223> Von Heijne matrix
     score 4.30000019073486
     seq LLYAHLSFTSKRA/VV

<221> polyA_site
<222> 1150..1161
```

```
<400> 331
aagatatcac tgtcttgttt tcacttagat cctacttaca aagtgagggt tattaacaga    60
ataaagcctt cctttaaagc tttataataa tcatatttat taataatgct gttgtgcata   120
cttatagtat gcatatattc agcatatgtt gcatgtsttc agaattacat aagatgaaat   180
ccctttcatt gcaacttgca agtgagaaaa gatccttagt ggctctggtg gaagaaatag   240
tatttcttct tctcagggtg tctccctgcc ttggcccctc ccagaagccc cggctttaaa   300
agtgaaaatg tttgaaacat gaaacatgtc tgtaggaagc atcagcatgg ccataagtgc   360
artgattttc atatatgcct ctgcccattt caaatatatt tttgacatga ataaatctaa   420
cagtatacar aataattcat gtaaraccct aacgtgtaca tgtgaaaaag catttctata   480
taatgtgagg agcactggcc atcaattagg gaaataaagg tcatgtaata ttgcaaattt   540
tcaaaataga gcsstgcaag ataactgcaa tcataccaaa aactatttga gtaaatggat   600
ttttaaagta atttttgttt aaaaaaattt atatttcaga agsagaaaat gtcaaatgat   660
agtctttgta a atg gtg gtg cac ctt ctc tat gca cat ctg tct ttt aca   710
            Met Val Val His Leu Leu Tyr Ala His Leu Ser Phe Thr
                -15                  -10                    -5
tca aaa aga gct gtg gtc atg cta aaa tta gag ata act ttt           752
Ser Lys Arg Ala Val Val Met Leu Lys Leu Glu Ile Thr Phe
            1           5                    10
tgaatgactt ggtcaagctg tgtgtaaaat atttaaccat aagtcaagta cagtgtacta   812
tgtttaataa agttacattt aatgcattta ttgcatatat gaatatatac atgaagaggc   872
tttatgtctt ctggtatttg attttgaatg tttttttaagt cagtggtgcc tttaggcaag   932
aactttcgaa attaatcatt ctttgtgttt tctgattttt caggtaacat gtacactatt   992
tagaaaccat catagtttat tcaccttaaa aaattgattg tattatttaa atatatcact  1052
tagatgggca tttcctataa ttaggatatt ccaaatagtt gctgaaatca attgtgccat  1112
tgaccaatgg atgcacttgg ttagccttaa ttttttyaaa aaaaaaaaa              1161
```

```
<210> 332
<211> 363
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 57..311

<221> sig_peptide
<222> 57..128
<223> Von Heijne matrix
     score 5.30000019073486
     seq LFHLLFLPHYIET/FK

<221> polyA_signal
<222> 332..337

<221> polyA_site
<222> 351..363
```

```
<400> 332
acatttctta ctgccttacg ctcatcctga ggtccacctt ggtctctaaa aacacc atg    59
                                                             Met
```

```
tgt tct cat gcc tcc atg tct ttt cac aca ctg ttc cat ttg ctc ttc        107
Cys Ser His Ala Ser Met Ser Phe His Thr Leu Phe His Leu Leu Phe
            -20                 -15                 -10
ctc cca cat tac att gaa act ttc aag cct cag tcg aaa cat tgc ttc        155
Leu Pro His Tyr Ile Glu Thr Phe Lys Pro Gln Ser Lys His Cys Phe
         -5                  1                   5
ttc tgg ata gca gcc ttc ttg aca tcc ctc ctc act ccc cag tcc cta        203
Phe Trp Ile Ala Ala Phe Leu Thr Ser Leu Leu Thr Pro Gln Ser Leu
10                  15                  20                  25
cag ggc ttc cat agc tct tta tgt gca ctt cga tcc cag cat ttt cca        251
Gln Gly Phe His Ser Ser Leu Cys Ala Leu Arg Ser Gln His Phe Pro
            30                  35                  40
tcg act tgt aat tgt ttc tgc tac ctg aca atc atc gcc ttg drd tac        299
Ser Thr Cys Asn Cys Phe Cys Tyr Leu Thr Ile Ile Ala Leu Xaa Tyr
            45                  50                  55
tgg gac aac ctt tgattactca ttatatcctc aataaatatt tgttgaacca           351
Trp Asp Asn Leu
            60
aaaaaaaaaa aa                                                         363
```

```
<210> 333
<211> 645
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 80..232

<221> sig_peptide
<222> 80..127
<223> Von Heijne matrix
      score 3.70000004768372
      seq IALTLIPSMLSRA/AG

<221> polyA_signal
<222> 617..622

<221> polyA_site
<222> 634..645

<400> 333
accttcttgt tatttatgct attctctttg tggctccatt cttctttcaa tcttctcagc     60
ttataaccgt ctttcccctt atg cta agg ata gcc ctt aca ctc atc cca tct    112
                        Met Leu Arg Ile Ala Leu Thr Leu Ile Pro Ser
                            -15                 -10
atg ctg tca agg gct gct ggt tgg tgc tgg tac aag gag ccc act cag      160
Met Leu Ser Arg Ala Ala Gly Trp Cys Trp Tyr Lys Glu Pro Thr Gln
-5                   1              5                   10
cag ttt tct tac ctt tgc ctg ccc tgc ctt tca tgg aat aar aaa ggc      208
Gln Phe Ser Tyr Leu Cys Leu Pro Cys Leu Ser Trp Asn Lys Lys Gly
            15                  20                  25
aac gtt ttg cag ctt cca aat ttc tgaaraaact aatctcarat tggcagttaa     262
Asn Val Leu Gln Leu Pro Asn Phe
            30                  35
agtcaaaatg ttgccaaata ttattccctt ttgcctaakt ttggctaccc ggttcaattg    322
ctttttattt ttaatgtctt gactcttcar agttcgtacc tcaaaaraac aatgaraaca    382
tttgctttgc tttctgctga atccctaatc tcaacaatct atacctggac tgtccagttc    442
tcctcctgtg ctatcttctc ttctatccaa gtaaatgta ygccaggarc tccttccctc    502
tarcaatttc tactaaaatg tccaagtara atgtttcctt ttacaatcaa attactgtat    562
```

ttattaattt gctaraatcc aktaaatcat tttggtagct ctggctgtgc tatcaataaa   622
aagatgaaag caaaaaaaaa aaa   645

<210> 334
<211> 400
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 91..291

<221> sig_peptide·
<222> 91..219
<223> Von Heijne matrix
      score 3.79999995231628
      seq LISVLYLIPKTLT/TN

<221> polyA_signal
<222> 367..372

<221> polyA_site
<222> 389..400

<400> 334
aacaaaagga gagttttata attcacttta aaaggagatt tgatggtaaa gtttaaagat   60
taaaatattt tgttcttcaa ttacagagcg atg acc cca cag tat ctg cct cac   114
                                    Met Thr Pro Gln Tyr Leu Pro His
                                                                   -40
ggt gga aaa tac caa gtt ctt gga gat tac tct ttg gca gtg gtc ttc   162
Gly Gly Lys Tyr Gln Val Leu Gly Asp Tyr Ser Leu Ala Val Val Phe
-35   -30   -25   -20
ccc ctg cac ttt tct gat cta att tct gtt tta tac ctt ata ccc aaa   210
Pro Leu His Phe Ser Asp Leu Ile Ser Val Leu Tyr Leu Ile Pro Lys
                -15   -10   -5
aca ctt act acc aac aca gct gtt aaa cat tct ata caa aaa aat tgt   258
Thr Leu Thr Thr Asn Thr Ala Val Lys His Ser Ile Gln Lys Asn Cys
                1   5   10
atg mat ctg gta tta gga aaa tta ctt tca cag taaatatcaa agaaaaaaga   311
Met Xaa Leu Val Leu Gly Lys Leu Leu Ser Gln
    15   20
ttaagggtct ctttgccatg cttttcatca tatgcaccaa atgtaaattt tgtacaataa   371
aattttattt cctaagyaaa aaaaaaaaa   400

<210> 335
<211> 496
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 196..384

<221> sig_peptide
<222> 196..240
<223> Von Heijne matrix
      score 6.69999980926514
      seq ILSTVTALTFARA/LD

<221> polyA_signal
<222> 461..466

<221> polyA_site
<222> 485..496

<400> 335

```
aaaaaaattgg tcccagtttt caccctgccg cagggctggc tggggagggc agcggtttag      60
attagccgtg gcctaggccg tttaacgggg tgacacgagc htgcagggcc gagtccaagg     120
cccggagata ggaccaaccg tcaggaatgc gaggaatgtt tttcttcgga ctctatcgag     180
gcacacagac agacc atg ggg att ctg tct aca gtg aca gcc tta aca ttt     231
              Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe
              -15             -10               -5
gcc aga gcc ctg gac ggc tgc aga aat ggc att gcc cac cct gca agt     279
Ala Arg Ala Leu Asp Gly Cys Arg Asn Gly Ile Ala His Pro Ala Ser
          .1              5                  10
gag aag cac aga ctc gag aaa tgt agg gaa ctc gag agc agc cac tcg     327
Glu Lys His Arg Leu Glu Lys Cys Arg Glu Leu Glu Ser Ser His Ser
    15               .20                 25
gcc cca gga tca acc cag cac cga aga aaa aca acc aga aga aat tat     375
Ala Pro Gly Ser Thr Gln His Arg Arg Lys Thr Thr Arg Arg Asn Tyr
30               35                 40                 45
tct tca gcc tgaaatgaak ccgggatcaa atggttgctg atcaragccc           424
Ser Ser Ala
atatttaaat tggaaaagtc aaattgasca ttattaaata aagcttgttt aatatgtctc   484
aaacaaaaaa aa                                                      496
```

<210> 336
<211> 968
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 54..590

<221> sig_peptide
<222> 54..227
<223> Von Heijne matrix
      score 3.5
      seq GGILMGSFQGTIA/GQ

<221> polyA_site
<222> 953..965

<400> 336

```
atatttgccc cttactttat cttgtgcctt gagaaattgc tggggagaga ggt atg       56
                                                          Met
tcc act ggg cag ctg tac agg atg gag gat ata ggg cgt ttc cac tcc     104
Ser Thr Gly Gln Leu Tyr Arg Met Glu Asp Ile Gly Arg Phe His Ser
        -55               -50                 -45
cag cag cca ggt tcc ctc acc cca agc tca ccc act gtt ggg gag att     152
Gln Gln Pro Gly Ser Leu Thr Pro Ser Ser Pro Thr Val Gly Glu Ile
    -40               -35                 -30
atc tac aat aac acc aga aac aca ttg ggg tgg att ggg ggt atc ctt     200
Ile Tyr Asn Asn Thr Arg Asn Thr Leu Gly Trp Ile Gly Gly Ile Leu
-25             .  -20                 -15                 -10
atg ggt tct ttt cag gga acc att gct gga caa ggc aca gga gcc acc     248
Met Gly Ser Phe Gln Gly Thr Ile Ala Gly Gln Gly Thr Gly Ala Thr
```

```
                    -5                     1              5
tcc att tct gag ctc tgc aag gga caa gaa cta gag cca tca ggg gct      296
Ser Ile Ser Glu Leu Cys Lys Gly Gln Glu Leu Glu Pro Ser Gly Ala
            10              15              20
ggg ctc act gtg gcc cca ccc caa gcc gtc agc ctc cag ggw atc tac      344
Gly Leu Thr Val Ala Pro Pro Gln Ala Val Ser Leu Gln Gly Ile Tyr
        25              30              35
acc ctg cct tgg ctg cta cag ctt ttt cac tcc act gcc cta rgg gna      392
Thr Leu Pro Trp Leu Leu Gln Leu Phe His Ser Thr Ala Leu Xaa Xaa
40              45·             50              55
dtt cag caa cct aat gga tct cta tct ctg aac atc tct tca tcc cat      440
Xaa Gln Gln Pro Asn Gly Ser Leu Ser Leu Asn Ile Ser Ser Ser His
                60              65              70
gct ccr rgt cca rca acc tgc acc ctg gaa cca gga gtg gac cct acc      488
Ala Pro Xaa Pro Xaa Thr Cys Thr Leu Glu Pro Gly Val Asp Pro Thr
                75              80              85
cga gct gtc tgt att aat ccc cat ccc cca cca cca atc tta aaa abc      536
Arg Xaa Val Cys Ile Asn Pro His Pro Pro Pro Pro Ile Leu Lys Xaa
        90              95              100
cct ctg tcc ccc tac cct aaa ccc cag tta ggt acc cat gct ggg caa      584
Pro Leu Ser Pro Tyr Pro Lys Pro Gln Leu Gly Thr His Ala Gly Gln
        105             110             115
gtc aat taacaattta tgcacaggta ctagttttat tgtattaccg ttccagggta      640
Val Asn
120
gctttgaaaa aagtatctca aaaaggcaac atgggccgag cgcagtggct cacgcctgta      700
atcccagcac tttgggaggc caaggtgggc agatcgcctg aggtctggag ttcaagacca      760
gcctggccaa cagggtgaaa ccccgtctct acaaaaatar gaaaattrgc caggtgtggt      820
ggcagacgtc tgtrgtccca gctattcagg agactgaggc acgagaattc catgaaccca      880
ggatgcggag gttgcagtga gccgagattg tgccactgcg ctccagcctg ggcgacagag      940
tggtattctg tttcaaaaaa aaaaamcm                                        968
```

```
<210> 337
<211> 901
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 133..846

<221> sig_peptide
<222> 133..345
<223> Von Heijne matrix
      score 9.39999961853027
      seq VVSFLLLLAGLIA/TY

<221> polyA_site
<222> 890..901

<400> 337
aagcagcttc caggatcctg agatccggag cagccggggt cggagcggct cctcaagagt      60
tactgatcta tnnatggcag agaaaaaaaa attgtgacca gagacgtgta gcaatgaaca     120
aggaacrtca ta atg rwn nnk ttc aca gac ccc tct tca gtg aat gaa aag     171
              Met Xaa Xaa Phe Thr Asp Pro Ser Ser Val Asn Glu Lys
                  -70             -65             -60
aag agg agg gag cgg gaa gaa agg cag aat att gtc ctg tgg aga cag      219
Lys Arg Arg Glu Arg Glu Glu Arg Gln Asn Ile Val Leu Trp Arg Gln
        -55             -50             -45
ccg ctc att acc ttg cag tat ttt tct ctg gaa atc ctt gta atc ttg      267
```

```
                  Pro Leu Ile Thr Leu Gln Tyr Phe Ser Leu Glu Ile Leu Val Ile Leu
                      -40                 -35                     -30
aag gaa tgg acc tca aaa tta tgg cat cgt caa agc att gtg gtg tct                    315
Lys Glu Trp Thr Ser Lys Leu Trp His Arg Gln Ser Ile Val Val Ser
    -25                 -20                 -15
ttt tta ctg ctg ctt gct ggg ctt ata gct acg tat tat gtt gaa gga                    363
Phe Leu Leu Leu Leu Ala Gly Leu Ile Ala Thr Tyr Tyr Val Glu Gly
-10                 -5                  1                   5
gtg cat caa cag tat gtg caa cgt ata gag aaa cag ttt ctt ttg tat                    411
Val His Gln Gln Tyr Val Gln Arg Ile Glu Lys Gln Phe Leu Leu Tyr
                10                  15                  20
gcc tac tgg ata ggc tta gga att ttg tct tct gtt ggg ctt gga aca                    459
Ala Tyr Trp Ile Gly Leu Gly Ile Leu Ser Ser Val Gly Leu Gly Thr
                25                  30                  35
ggg ctg cac acc ttt ctg ctt tat ctg ggt cca cat ata gcc tca gtt                    507
Gly Leu His Thr Phe Leu Leu Tyr Leu Gly Pro His Ile Ala Ser Val
            40                  45                  50
aca tta gct gct tat gaa tgc aat tca gtt aat ttt ccc gaa cca ccc                    555
Thr Leu Ala Ala Tyr Glu Cys Asn Ser Val Asn Phe Pro Glu Pro Pro
55                  60                  65                  70
tat cct gat cag att att tgt cca gat gaa gag ggc act gaa gga acc                    603
Tyr Pro Asp Gln Ile Ile Cys Pro Asp Glu Glu Gly Thr Glu Gly Thr
                75                  80                  85
att tct ttg tgg agt atc atc tca aaa gtt agg att gaa gcc tgc atg                    651
Ile Ser Leu Trp Ser Ile Ile Ser Lys Val Arg Ile Glu Ala Cys Met
                90                  95                  100
tgg ggt atc ggt aca gca atc gga gag ctg cct cca tat ttc atg gcc                    699
Trp Gly Ile Gly Thr Ala Ile Gly Glu Leu Pro Pro Tyr Phe Met Ala
                105                 110                 115
aga gca gct cgc ctc tca ggt gct gaa cca gat gat gaa gag tat cag                    747
Arg Ala Ala Arg Leu Ser Gly Ala Glu Pro Asp Asp Glu Glu Tyr Gln
    120                 125                 130
gaa ttt gaa gag atg ctg gaa cat gca gag tct gca caa gta aga aca                    795
Glu Phe Glu Glu Met Leu Glu His Ala Glu Ser Ala Gln Val Arg Thr
135                 140                 145                 150
gtg ggg ata gaa aat aga aca ctt tac ttc ttc cta aag agg cta tta                    843
Val Gly Ile Glu Asn Arg Thr Leu Tyr Phe Phe Leu Lys Arg Leu Leu
                155                 160                 165
agg taaaattgtt agtagttact ctgaagaaga aaactgctaa agtaaaaaaa aaaaa           901
Arg
```

```
<210> 338
<211> 1347
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 138..671

<221> sig_peptide
<222> 138..248
<223> Von Heijne matrix
      score 3.5
      seq LVFNFLLILTILT/IW

<221> polyA_signal
<222> 1319..1324

<221> polyA_site
```

<222> 1338..1347

<400> 338
aagaatgctt gtgaagtagc aactaaagtg gcagtgtttc ttctgaaatt ctcaggcagt   60
cagactgtct taggcaaatc ttgataaaat agcccttatc caggttttta tctaaggaat   120
cccaagaaga ctggggga atg gag aga cag tca agg gtt atg tca gaa aag   170
                     Met Glu Arg Gln Ser Arg Val Met Ser Glu Lys
                      -35   -30
gat gag tat cag ttt caa cat cag gga gcg gtg gag ctg ctt gtc ttc   218
Asp Glu Tyr Gln Phe Gln His Gln Gly Ala Val Glu Leu Leu Val Phe
     -25   -20   -15
aat ttt ttg ctc atc ctt acc att ttg aca atc tgg tta ttt aaa aat   266
Asn Phe Leu Leu Ile Leu Thr Ile Leu Thr Ile Trp Leu Phe Lys Asn
-10   -5   1   5
cat cga ttc cgc ttc ttg cat gaa act gga gga gca atg gtg tat ggc   314
His Arg Phe Arg Phe Leu His Glu Thr Gly Gly Ala Met Val Tyr Gly
             10   15   20
ctt aya atg gga cta att tta csa tat gct aca gca cca act gat att   362
Leu Xaa Met Gly Leu Ile Leu Xaa Tyr Ala Thr Ala Pro Thr Asp Ile
             25   30   35
gaa agt ggr rct gtc tat gac tgt gta aaa cta act ttc agt cca tca   410
Glu Ser Gly Xaa Val Tyr Asp Cys Val Lys Leu Thr Phe Ser Pro Ser
     40   45   50
act ctg ctg gtt aat atc act gac caa gtt tat gar tat aaa tac aar   458
Thr Leu Leu Val Asn Ile Thr Asp Gln Val Tyr Glu Tyr Lys Tyr Lys
55   60   65   70
aga gaa ata agt cag cac amc atc aat cct cat cam gga aat gct ata   506
Arg Glu Ile Ser Gln His Xaa Ile Asn Pro His Xaa Gly Asn Ala Ile
                 75   80   85
ctt gaa aag atg aca ttt gat cca raa atc ttc ttc aat gtt tta ctg   554
Leu Glu Lys Met Thr Phe Asp Pro Xaa Ile Phe Phe Asn Val Leu Leu
             90   95   100
cca cca att ata ttt cat gca gga tat agt cta aag aag aga cac ttt   602
Pro Pro Ile Ile Phe His Ala Gly Tyr Ser Leu Lys Lys Arg His Phe
             105   110   115
ttt caa aac tta gga tct att tta acg tat gcc ttc ttg gga act gcc   650
Phe Gln Asn Leu Gly Ser Ile Leu Thr Tyr Ala Phe Leu Gly Thr Ala
     120   125   130
atc tcc tgc atc gtc ata ggg taagtgacat tcggagctca agttgcaggt   701
Ile Ser Cys Ile Val Ile Gly
135   140
ggctgtgggg tcygtgatct gtgtgaggga tctaacactt ccaggattct tgctggckgg   761
gaaaattgtc ttttttttar tawatcacaw atttgtatgt tttttcwgac ttaattccac   821
ggcttckgam aaatacaagg cttcaaatca aagcaaacta waggattgct ggactttctc   881
tgtgagttct ggacttctga cttagggaat gtggatcact tgccttgagt tatgtgaagc   941
gcattgcatt cttctttag tttgagtaat sccgatatgc tcactgcatt cttttttgtc   1001
ttgtattgag agaccttacc tgtatttggc aggagtgcaa aagtaactat atgccaagag   1061
tttctttct aaaggaaagt ttacaagaca gcagtctgaa acagatatgt ccaaatatca   1121
acagagttgc ttaatacagg gatagctttt cagttaatac cctgtagaat gcagactctt   1181
tttttcattg tatttcttg attatgctac tgagccctaa gtcacacgtt atatactctg   1241
gcttgcagct catcataaag taaaatgtgg taccaaatgg tgaaggcaat ccagcctctg   1301
ataatcccgt ccaatacatt aaagctccac tgcaggaaaa aaaaaa   1347

<210> 339
<211> 987
<212> DNA
<213> Homo sapiens

<220>
<221> CDS

<222> 124..411

<221> sig_peptide
<222> 124..186
<223> Von Heijne matrix
      score 6.30000019073486
      seq MVALCCCLWKISG/CE

<221> polyA_signal
<222> 948..953

<221> polyA_site
<222> 971..983

<400> 339
aagacgctgc ctttagggag agataaaaag cataatgaca ttagctagga aagttaattt   60
tcagttctta ctgaagtgct gtatgaaact gaaatttcca aggaactgaa ttttgtgagc   120
caa atg agc atg caa ttc ttg ttt aag atg gtg gcc tta tgc tgt tgt   168
    Met Ser Met Gln Phe Leu Phe Lys Met Val Ala Leu Cys Cys Cys
        -20   -15   -10
ctc tgg aag atc tcc ggc tgt gag gaa gtc cct cta act tac aac ctg   216
Leu Trp Lys Ile Ser Gly Cys Glu Glu Val Pro Leu Thr Tyr Asn Leu
    -5   1   5   10
ctc aag tgc ctc cta gat aaa gcg cac tgt gta ctc ctg aca cct tgt   264
Leu Lys Cys Leu Leu Asp Lys Ala His Cys Val Leu Leu Thr Pro Cys
            15   20   25
ggt tac atc ttt tcc ttg atc agt cca gaa att ctc aaa ctc act tta   312
Gly Tyr Ile Phe Ser Leu Ile Ser Pro Glu Ile Leu Lys Leu Thr Leu
            30   35   40
atc act ttg cav atc ctc tta ata ctc aaa aat cta cac tta ctg tgg   360
Ile Thr Leu Xaa Ile Leu Leu Ile Leu Lys Asn Leu His Leu Leu Trp
            45   50   55
ctg aca gtt tca agc awa tgt gtt cat cgc agt agt gca aga aaa gaa   408
Leu Thr Val Ser Ser Xaa Cys Val His Arg Ser Ser Ala Arg Lys Glu
    60   65   70
aag tagaagaacc ctgcagagat ttgatggaac ccagcttcta ttcattaaaa   461
Lys
75
ccaatggcaa aatataaagc aaataggagg tgacgaaggt tacaaaaata cgtattgttt   521
atgttttccc tggggtgtgc tgattgtcag gcatcagttc cctgtgccat tcattcccca   581
acacagcatg catcagaaat tttatcaata aatgctttct ctctcaatgt tcaacctatg   641
ctgatagacc attaaataca gtttttgggt tcacagcttg tcatcatcat ttgtctatac   701
ctgtggcaaa gaatatctaa taagatactc tcagcatttt gcacacttaa actaagatgc   761
tgaatgctgt attttacgga ataatcagcc acattaaatt tggagactca acaagcatgc   821
tgtgaacatt caacattagg tttaaatttt atttttaaaa gttaataata aaaggatata   881
tgttaagtat tatgaaaccc tgcatatact gtaataaaat ggtggatgtg aatggacaat   941
atatgcaata aaatttataa tttgattcya aaaaaaaaaa aamccv   987

<210> 340
<211> 748
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 372..494

<221> sig_peptide
<222> 372..443
<223> Von Heijne matrix

```
        score 5.30000019073486
        seq RILLLHFYCLLRS/SE

<221> polyA_signal
<222> 708..713

<221> polyA_site
<222> 732..745

<400> 340
acatgaaatg tgcttggtct gtgatctctt ggtcagatat ctgccttcca ggcgatcctt       60
tgaggttgtg taattcagct ggccctggct cctggtccct gttactgagc tgggcagtcg      120
aaccgaaggc agatgagctc aagatcatgc cttgggaagc atggtgctct aggggtgcct      180
ttttattcct ttcattgtat tatagactgt ttccaagttt atggttagaa atggtaaagt      240
gggtctggtg ttttgaggta gaacccagcc tagggcaaga tatgaactgt tcttgaggta      300
gaaatgtcta cagtcagttg tttcatctag cttgcatctt aaaacacaaa cccttcagtt      360
gctttcactt a atg cac aca ttt gcc aat gac aga ggg tta tac agg atc       410
            Met His Thr Phe Ala Asn Asp Arg Gly Leu Tyr Arg Ile
                       -20                     -15

ctt ctt tta cat ttc tat tgt ctg cta cgc tca tca gag tat att ttg       458
Leu Leu Leu His Phe Tyr Cys Leu Leu Arg Ser Ser Glu Tyr Ile Leu
    -10                 -5              1              5

ggg tac aag gtt ttg ggg gtt ttt tty ccc att ttg taactgcctt           504
Gly Tyr Lys Val Leu Gly Val Phe Phe Pro Ile Leu
            10                  15

attgaaaadt aaktgccctt ccattccagg cctcctcata ttgtacttgt ttcctgccaa      564
atctggggga tcatttgtat tttaactttg taatctatgg ctctgtactg ttgaaagstc      624
tcaattctgt ggggtctcct tagtatgtat gtgacttttc atgttgcaat atcacacgat      684
gggatggccc gacttttgct cttaataaat aatctgaatg agtaagaraa aaaaaaaaa      744
accc                                                                    748
```

```
<210> 341
<211> 1106
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 112..450

<221> sig_peptide
<222> 112..192
<223> Von Heijne matrix
        score 7.19999980926514
        seq SLLFFLLLEGGXT/EQ

<221> polyA_signal
<222> 1053..1058

<221> polyA_site
<222> 1095..1106

<400> 341
aagacctcgg aacgagagcg ccccggggag ctcggagcgc gtgcacgcgt ggcavacgga       60
gaaggcvakk rcnnnnrctt gaaggttctg tcacctttg cagtggtcca a atg aga      117
                                                        Met Arg

raa aag tgg aaa atg gga ggc atg aaa tac atc ttt tcg ttg ttg ttc       165
Xaa Lys Trp Lys Met Gly Gly Met Lys Tyr Ile Phe Ser Leu Leu Phe
-25                 -20               -15               -10

ttt ctt ttg cta gaa gga ggc kaa aca gag caa gtr amn cat tca gag       213
```

```
Phe Leu Leu Leu Glu Gly Gly Xaa Thr Glu Gln Val Xaa His Ser Glu
            -5                  1               5
aca tat tgc atg ttt caa gac aag aag tac aga gtg ggt gag aga tgg        261
Thr Tyr Cys Met Phe Gln Asp Lys Lys Tyr Arg Val Gly Glu Arg Trp
        10                  15                  20
cat cct tac ctg gaa cct tat ggg ttg gtt tac tgc gtg aac tgc atc      309
His Pro Tyr Leu Glu Pro Tyr Gly Leu Val Tyr Cys Val Asn Cys Ile
    25                  30                  35
tgc tca gag aat ggg aat gtg ctt tgc agc cga gtc aga tgt cca aat      357
Cys Ser Glu Asn Gly Asn Val Leu Cys Ser Arg Val Arg Cys Pro Asn
40                  45                  50                  55
gtt cat tgc ctt tct cct gtg cat att cct cat ctg tgc tgc cct cgc      405
Val His Cys Leu Ser Pro Val His Ile Pro His Leu Cys Cys Pro Arg
                60                  65                  70
tgc cca gaa gac tcc tta ccc cca gtg aac aat rwg gtg acc agc          450
Cys Pro Glu Asp Ser Leu Pro Pro Val Asn Asn Xaa Val Thr Ser
            ·75                 80                  85
tagtcttgck agtacaatgg gacaacttac caacatggas agctgttcgt agct·rrggg      510
ctctttcaga atcggcaacc cmatcaatgc acccagtgca gctgttcgga rggaaacktg      570
tattgtggtc tcaagacttg ccccaaatta acctgtgcct tccagtctc tgttccarat       630
tcctgctgcc gggtwtgcag argagatgga caactgtcat gggaacmttc tgatggtgat      690
atcttccggc aacctgccaa cagagaagca agacattctt accaccgctc tcactatgat      750
cctccaccaa gccgacaggc tggaggtctg tcccgctttc ctggggccag aagtcaccgg      810
ggagctctta tggattccca gcaagcatca ggaaccattg tgcaaattgt catcaataac      870
aaacacaagc atggacaagt gtgtgtttcc aatggaaaga cctattctca tggcgagtcc      930
tggcacccaa acctccgggc atttggcatt gtggagtgtg tgctatgtac ttgtaatgtc      990
accaagcaag agtgtaagaa aatccactgc cccaatcgat accctgcaa gtatcctcaa      1050
aaaatagacg gaaaatgctg caaggtgtgt ccaggtaaaa aagcaaaaaa aaaaaa        1106
```

```
<210> 342
<211> 1191
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 117..866

<221> sig_peptide
<222> 117..170
<223> Von Heijne matrix
      score 10.6999998092651
      seq LILLALATGLVGG/ET

<221> polyA_signal
<222> 1159..1164

<221> polyA_site
<222> 1178..1190

<400> 342
aaaacccagc ctacctgctg tagctgccgc cactgccgtc tccgccgcca ctggwccccc      60
agagcbnmag ccccagagcc taggaacctg gggcccgctc ctcccccctc caggcc atg     119
                                                               Met
agg att ctg cag tta atc ctg ctt gct ctg gca aca ggg ctt gta ggg      167
Arg Ile Leu Gln Leu Ile Leu Leu Ala Leu Ala Thr Gly Leu Val Gly
        -15                 -10                 -5
gga gag acc agg atc atc aag ggg ttc gag tgc aag cct cac tcc cag      215
Gly Glu Thr Arg Ile Ile Lys Gly Phe Glu Cys Lys Pro His Ser Gln
    1                   5                   10                  15
```

```
ccc tgg cag gca gcc ctg ttc gag aag acg cgg cta ctc tgt ggg gcg        263
Pro Trp Gln Ala Ala Leu Phe Glu Lys Thr Arg Leu Leu Cys Gly Ala
            20                  25              30
acg ctc atc gcc ccc aga tgg ctc ctg aca gca gcc cac tgc ctc aag        311
Thr Leu Ile Ala Pro Arg Trp Leu Leu Thr Ala Ala His Cys Leu Lys
            35                  40              45
ccc cgc tac ata ktt cac ctg ggg cag cac aac ctc cag aag gag gag        359
Pro Arg Tyr Ile Xaa His Leu Gly Gln His Asn Leu Gln Lys Glu Glu
            50                  55              60
ggc tgt gag car acc cgg aca gcc act gag tcc ttc ccc cac ccc ggc        407
Gly Cys Glu Gln Thr Arg Thr Ala Thr Glu Ser Phe Pro His Pro Gly
            65                  70              75
ttc aac aac agc ctc ccc aac aaa gac cam mgc aat gac atc atg ctg        455
Phe Asn Asn Ser Leu Pro Asn Lys Asp Xaa Xaa Asn Asp Ile Met Leu
80              85                  90                  95
gtg aak atg gma tcg cca gtc tcc atc acc tgg gct gtg cga ccc ctc        503
Val Xaa Met Xaa Ser Pro Val Ser Ile Thr Trp Ala Val Arg Pro Leu
                100                 105                 110
acc ctc tcc tca cgc tgt gtc act gct ggc acc agc tgc ctc att tcc        551
Thr Leu Ser Ser Arg Cys Val Thr Ala Gly Thr Ser Cys Leu Ile Ser
                115                 120                 125
ggc tgg ggc agc acg tcc agc ccc cag tta cgc ctg cct cac acc ttg        599
Gly Trp Gly Ser Thr Ser Ser Pro Gln Leu Arg Leu Pro His Thr Leu
            130                 135                 140
cga tgc gcc aac atc acc atc att gag cac cag aag tgt gag aac gcc        647
Arg Cys Ala Asn Ile Thr Ile Ile Glu His Gln Lys Cys Glu Asn Ala
            145                 150                 155
tac ccc ggc aac atc aca gac acc atg gtg tgt gcc agc gtg cag gaa        695
Tyr Pro Gly Asn Ile Thr Asp Thr Met Val Cys Ala Ser Val Gln Glu
160             165                 170                 175
ggg ggc aag gac tcc tgc cag ggt gac tcc ggg ggc cct ctg gtc tgt        743
Gly Gly Lys Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys
                180                 185                 190
aac cag tct ctt caa ggc att atc tcc tgg ggc cag gat ccg tgt gcg        791
Asn Gln Ser Leu Gln Gly Ile Ile Ser Trp Gly Gln Asp Pro Cys Ala
            195                 200                 205
atc acc cga aag cct ggt gtc tac acg aaa gtc tgc aaa tat gtg gac        839
Ile Thr Arg Lys Pro Gly Val Tyr Thr Lys Val Cys Lys Tyr Val Asp
            210                 215                 220
tgg atc cag gag acg atg aag aac aat tagactggac ccacccacca            886
Trp Ile Gln Glu Thr Met Lys Asn Asn
            225                 230
cagcccatca ccctccattt ccacttggtg tttggttcct gttcactctg ttaataagaa    946
accctaagcc aagaccctct acgaacattc tttgggcctc ctggactaca ggagatgctg    1006
tcacttaata atcaacctgg ggttcgaaat cagtgagacc tggattcaaa ttctgccttg    1066
aaatattgtg actctgggaa tgacaacacc tggtttgttc tctgttgtat ccccagcccc    1126
aaakwcagct cctggccata tatcaaggtt tcaataaata tttgctaaat gaawaaaaaa    1186
aaaac                                                               1191
```

```
<210> 343
<211> 1070
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 13..465

<221> sig_peptide
<222> 13..75
```

<223> Von Heijne matrix
      score 3.90000009536743
      seq PVAVTAAVAPVLS/IN

<221> polyA_signal
<222> 1035..1040

<221> polyA_site
<222> 1060..1070

<400> 343
```
agagtcggga aa atg gct gcg agt acc tcc atg gtc ccg gtg gct gtg acg        51
              Met Ala Ala Ser Thr Ser Met Val Pro Val Ala Val Thr
                -20                 -15                 -10
gcg gca gtg gcg cct gtc ctg tcc ata aac agc gat ttc tca gat ttg         99
Ala Ala Val Ala Pro Val Leu Ser Ile Asn Ser Asp Phe Ser Asp Leu
        -5                   1                   5
cgg gaa att aaa aag caa ctg ctg ctt att gcg ggc ctt acc cgg gag        147
Arg Glu Ile Lys Lys Gln Leu Leu Leu Ile Ala Gly Leu Thr Arg Glu
    10                  15                  20
cgg ggc cta cta cac agt agc aaa tgg tcg gcg gag ttg gct ttc tct        195
Arg Gly Leu Leu His Ser Ser Lys Trp Ser Ala Glu Leu Ala Phe Ser
25                  30                  35                  40
ctc cct gca ttg cct ctg gcc gag ctg caa ccg cct ccg cct att aca        243
Leu Pro Ala Leu Pro Leu Ala Glu Leu Gln Pro Pro Pro Pro Ile Thr
                45                  50                  55
gag gaa gat gcc cag gat atg gat gcc tat acc ctg gcc aag gcc tac        291
Glu Glu Asp Ala Gln Asp Met Asp Ala Tyr Thr Leu Ala Lys Ala Tyr
            60                  65                  70
ttt gac gtt aaa gag tat gat cgg gca gca cat ttc ctg cat ggc tgc        339
Phe Asp Val Lys Glu Tyr Asp Arg Ala Ala His Phe Leu His Gly Cys
        75                  80                  85
aat gca aga aaa gcc tat ttt ctg tat atg tat tcc aga tat ctg gtg        387
Asn Ala Arg Lys Ala Tyr Phe Leu Tyr Met Tyr Ser Arg Tyr Leu Val
    90                  95                  100
agg gcc att tta aaa tgt cat tct gcc ttt agt gaa aca tcc ata ttt        435
Arg Ala Ile Leu Lys Cys His Ser Ala Phe Ser Glu Thr Ser Ile Phe
105                 110                 115                 120
aga acc aat gga aaa gtt aaa tct ttt aaa tagcttagca gtgggccact        485
Arg Thr Asn Gly Lys Val Lys Ser Phe Lys
                125                 130
gaatgaatgt actttataca tagcaataat aaaaaaaaga tatcataaat aaagttaaaa    545
aggatggtaa aaaaaaaat attcttagga atgactaaca ggataagtaa caacctgatt    605
atttatttac tttaggttat ataaggttct tcatgcctgt gaattaatat tattgtgtaa   665
gaattaagtt aaaaagcctg ggctgacttt taaatttata aattcattta tcatgtttat   725
agtatattta ttgttttct ttcatggcta ttaaaaagta tgactgtaaa ggacaatgca    785
agtaaaccaa cttaatactg tattgaataa taagtacaat ttattatttt actttgaaac   845
attatgaatt tactttccta cttttttctta gttgttatct atataaattg attaaaaaaa  905
cattttatgt acttctcatt tcctagtaca ggttgagtat cccttatttg aagtgcttgg   965
gaccaaaagt gtttcagatt tcagattttt ttcagatttt ggtatatttg cattatactt  1025
actggttgaa ataaaaaatg ctgcagtgag tgtcaaaaaa aaaaa                   1070
```

<210> 344
<211> 1213
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 2..718

```
<221> sig_peptide
<222> 2..76
<223> Von Heijne matrix
      score 3.90000009536743
      seq RVGLLLGGGGVYG/SR

<221> polyA_signal
<222> 1170..1175

<221> polyA_site
<222> 1203..1213


<400> 344
a atg ccc cgg aag cgg aag tgc gat ctt cgg gct gtc aga gtt ggt ctg      49
  Met Pro Arg Lys Arg Lys Cys Asp Leu Arg Ala Val Arg Val Gly Leu
    -25             -20             -15             -10
tta ctc ggt ggt ggc gga gtc tac gga agc cgt ttt cgc ttc act ttt       97
Leu Leu Gly Gly Gly Gly Val Tyr Gly Ser Arg Phe Arg Phe Thr Phe
              -5              1               5
cct ggc tgt aga gcg ctt tcc ccc tgg cgg gtg aga vtg cag aga cga      145
Pro Gly Cys Arg Ala Leu Ser Pro Trp Arg Val Arg Xaa Gln Arg Arg
          10              15              20
agg tgc gag atg agc act atg ttc gcg gac act ctc ctc atc gtt ttt      193
Arg Cys Glu Met Ser Thr Met Phe Ala Asp Thr Leu Leu Ile Val Phe
      25              30              35
atc tct gtg tgc acg gct ctg ctc gca gag ggc ata acc tgg gtc ctg      241
Ile Ser Val Cys Thr Ala Leu Leu Ala Glu Gly Ile Thr Trp Val Leu
40              45              50              55
gtt tac agg aca gac aag tac aag aga ctg aag gca gaa gtg gaa aaa      289
Val Tyr Arg Thr Asp Lys Tyr Lys Arg Leu Lys Ala Glu Val Glu Lys
              60              65              70
cag agt aaa aaa ttg gaa aag aag aag gaa aca ata aca gag tca gct      337
Gln Ser Lys Lys Leu Glu Lys Lys Lys Glu Thr Ile Thr Glu Ser Ala
          75              80              85
ggt cga caa cag aaa aar aaa ata gag aga cdd kaa kas amc ctg arg      385
Gly Arg Gln Gln Lys Lys Lys Ile Glu Arg Xaa Xaa Xaa Xaa Leu Xaa
          90              95              100
aat aac aac aga gat cta tca atg gtt cga atg aaa tcc atg ttt gct      433
Asn Asn Asn Arg Asp Leu Ser Met Val Arg Met Lys Ser Met Phe Ala
      105             110             115
att ggc ttt tgt ttt act gcc cta atg gga atg ttc aat tcc ata ttt      481
Ile Gly Phe Cys Phe Thr Ala Leu Met Gly Met Phe Asn Ser Ile Phe
120             125             130             135
gat ggt aga gtg gtg gca aag ctt cct ttt acc cct ctt tct tas rtc      529
Asp Gly Arg Val Val Ala Lys Leu Pro Phe Thr Pro Leu Ser Xaa Xaa
              140             145             150
sra gga ctg tct cat cga aat ctg ctg gga gat gac acc aca gac tgt      577
Xaa Gly Leu Ser His Arg Asn Leu Leu Gly Asp Asp Thr Thr Asp Cys
          155             160             165
tcc ttc att ttc ctg taw att ctc tgt act atg tcg att cga cag aac      625
Ser Phe Ile Phe Leu Xaa Ile Leu Cys Thr Met Ser Ile Arg Gln Asn
          170             175             180
att cag aag att ctc ggc ctt gcc cct tca cga gcc gcc acc aag cag      673
Ile Gln Lys Ile Leu Gly Leu Ala Pro Ser Arg Ala Ala Thr Lys Gln
          185             190             195
gca ggt gga ttt ctt ggc cca cca cct cct tct ggg aag ttc tct         718
Ala Gly Gly Phe Leu Gly Pro Pro Pro Ser Gly Lys Phe Ser
200             205             210
tgaactcaag aactctttat tttctakcat tctttctaga cacacacaca tcagactggc    778
aactgttttg tascaagagc cataggtagc cttackactt gggcctcttt ctagttttga    838
attatttcta agccttttgg gtatkattag agtgaaaatg gcagccagca aacttgatag    898
```

```
tgctttggt  cctagatgat  ttttatcaaa  taagtggatt  gattagttaa  gttcaggtaa      958
tgtttatgta  atgaaaaaca  aatagcatcc  ttcttgtttc  atttacataa  gtattttctg     1018
tgggaccgac  tctcaaggca  ctgtgtatgc  cctgcaagtt  ggctgtctat  gagcatttag     1078
agatttagaa  gaaaaattta  gtttgtttaa  cccttgtaac  tgttgttttt  gttgttgttt     1138
tttttttcaag  ccaaatacat  gacataarat  caataaaarag  gccaaatttt  tasctgtttt    1198
atgtaaaaaa  aaaaa                                                          1213
```

```
<210> 345
<211> 978
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 86..709

<221> sig_peptide
<222> 86..361
<223> Von Heijne matrix
      score 6.30000019073486
      seq LLMSILALIFIMG/NS

<221> polyA_signal
<222> 943..948

<221> polyA_site
<222> 963..973

<400> 345
aaagcatcct  tccctaggac  tgctgtaagc  tttgagcctc  tagcaggaga  catgcctcgg      60
ggacgaaaga  gtcggcgccg  ccgta atg cga gag ccg cag aag aga acc gca         112
                             Met Arg Glu Pro Gln Lys Arg Thr Ala
                                        -90                 -85
aca atc gca aaa tyc rrg gcs tva gag ggc ctc cga gac ccc tat ggc           160
Thr Ile Ala Lys Xaa Xaa Ala Xaa Glu Gly Leu Arg Asp Pro Tyr Gly
            -80                 -75                 -70
cgc ctc tgt ggt agc gag cac ccc cga aga cca cct gag cgg ccc gag           208
Arg Leu Cys Gly Ser Glu His Pro Arg Arg Pro Pro Glu Arg Pro Glu
        -65                 -60                 -55
gaa gac ccg agc act cca gag gag gcc tct acc acc cct gaa gaa gcc           256
Glu Asp Pro Ser Thr Pro Glu Glu Ala Ser Thr Thr Pro Glu Glu Ala
    -50                 -45                 -40
tcg agc act gcc caa gca caa aag cct tca gtg ccc cgg agc aat ttt           304
Ser Ser Thr Ala Gln Ala Gln Lys Pro Ser Val Pro Arg Ser Asn Phe
-35                 -30                 -25                 -20
cag ggc acc aag aaa agt ctc ctg atg tct ata tta gcg ctc atc ttc           352
Gln Gly Thr Lys Lys Ser Leu Leu Met Ser Ile Leu Ala Leu Ile Phe
            -15                 -10                 -5
atc atg ggc aac agc gcc aag gaa gct ctg gtc tgg aaa gtg ctg ggg           400
Ile Met Gly Asn Ser Ala Lys Glu Ala Leu Val Trp Lys Val Leu Gly
                1                   5                   10
aag tta gga atg cag cct gga cgt cas cac agc atc ttt gga gat ccg           448
Lys Leu Gly Met Gln Pro Gly Arg Xaa His Ser Ile Phe Gly Asp Pro
            15                  20                  25
aag aar atc gtc aca gaa ran ttt gtg cgc aga ggg tac ctg att tat           496
Lys Lys Ile Val Thr Glu Xaa Phe Val Arg Arg Gly Tyr Leu Ile Tyr
30                  35                  40                  45
ara ccg gtg ccc cgt abc agt ccg gtg gag tat gas ttc ttc tgg ggg           544
Xaa Pro Val Pro Arg Xaa Ser Pro Val Glu Tyr Xaa Phe Phe Trp Gly
                50                  55                  60
```

```
ccc cga gca cac gtg gaa tcg agc ara ctg aaa stc wtg cat ttt gtg      592
Pro Arg Ala His Val Glu Ser Ser Xaa Leu Lys Xaa Xaa His Phe Val
         65                  70                  75
gca agg gtt cgt aac cga tgc tct aaa gac tgg cct tgt aat tat gac      640
Ala Arg Val Arg Asn Arg Cys Ser Lys Asp Trp Pro Cys Asn Tyr Asp
         80                  85                  90
tgg gat tcg gac gat gat gca gag gtt gag gct atc ctc aat tca ggt      688
Trp Asp Ser Asp Asp Asp Ala Glu Val Glu Ala Ile Leu Asn Ser Gly
         95                 100                 105
gct arg ggt tat tcc gcc cct taagtaratc tgaggcagac ccttggggt         739
Ala Xaa Gly Tyr Ser Ala Pro
110                 115
gtaaaagaga gtcacaggta ccccaaggag tagatgccag ggtcctaagt tgaaaatgmt   799
gtcgattggg ggcgggggac actgtatttg atatttgtga tcagtgatca ttgttcaact   859
gcgaaataga gtgtttgctt ttgataatgg aaaattgtat tcgttttaaa attccgtttg   919
ttgagaataa caatatgttt aaaaatataa ttgaacaaat tttaaaaaaa aaaamcccy    978
```

```
<210> 346
<211> 810
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 63..320

<221> sig_peptide
<222> 63..179
<223> Von Heijne matrix
      score 3.90000009536743
      seq VLAIGLLHIVLLS/IP

<221> polyA_signal
<222> 771..776

<221> polyA_site
<222> 799..810

<400> 346
agggaaccga tcccgggccg ttgatcttcg gccccacacg aacagcagag aggggcatca    60
gg atg aat gtk ggc aca gcg cac ags dag gtg aac ccc aac acg cgg      107
   Met Asn Val Gly Thr Ala His Xaa Xaa Val Asn Pro Asn Thr Arg
           -35                 -30                 -25
gtk atg aac agc cgt ggc atc tgg ctc tcc tac gtg ctg gcc atc ggt     155
Val Met Asn Ser Arg Gly Ile Trp Leu Ser Tyr Val Leu Ala Ile Gly
           -20                 -15                 -10
ctc ctc cac atc gtg ctg ctg agc atc ccg ttt gtk agt gtc cct gtc     203
Leu Leu His Ile Val Leu Leu Ser Ile Pro Phe Val Ser Val Pro Val
           -5                  1                   5
gtc tgg acc ctc acc aac ctc att cac aac atg ggc atg tat atc ttc     251
Val Trp Thr Leu Thr Asn Leu Ile His Asn Met Gly Met Tyr Ile Phe
     10                  15                  20
ctg cac acg gtg aag ggg aca ccc ttt gag acc ccg gac cag ggc aag     299
Leu His Thr Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys
25                  30                  35                  40
gcg agg ctg cta acc cac tgg tgagcagatg gattatgggg tccagttcac        350
Ala Arg Leu Leu Thr His Trp
                45
ggcctctcgg aakttcttga ccatcacacc catcgtgctg tacttcctca ccagcttcta   410
cactaaktac raccaaatcc attttgtgct caacaccgtg tccctgatra gcgtgcttat   470
```

```
ccccaagctg ccccagctcc acggaktccg gatttttgga atcaataakt actgaaaaktg    530
cascccccttc ccctgcccag ggtggcaggg gaggggtagg gtaaaaggca tktgctgcaa    590
chctgaaaac araaaraara rscctctgga cactgccara ratggggggtt gagcctctgg    650
cctaattttcc cccctcgctt ccccccagtag ccaacttgga gtagcttgta ytgggggttgg   710
ggtaggcccc ctgggctctg accttttctg aattttttga tctttttcctt ttgcttttttg    770
aatararact ccatggagtt ggtcatggaa aaaaaaaaaa                            810
```

<210> 347
<211> 771
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 299..418

<221> sig_peptide
<222> 299..379
<223> Von Heijne matrix
      score 3.59999990463257
      seq LTLLLITPSPSPL/LF

<221> polyA_signal
<222> 739..744

<221> polyA_site
<222> 762..771

<400> 347
```
accttgggct ccaaattcta gctcataaag atgcaagtkt tgcaatttcc tataaatggt     60
taagaaaaga gcaagctgtc cagagagtga gaagtttgaa aagagaggtg cataagagag    120
aaatgatgtc catttgagcc ccaccacgga ggttatgtgg tcccaaaagg aatgatggcc    180
aagcaattaa tttttcctcc tagttcttag cttgcttctg cattgattgg ctttacacaa    240
ctggcattta gtctgcatta cacaaataga cactaattta tttggaacaa gcagcaaa     298
atg aga act tta ttt ggt gca gtc agg gct cca ttt agt tcc ctc act     346
Met Arg Thr Leu Phe Gly Ala Val Arg Ala Pro Phe Ser Ser Leu Thr
        -25                 -20                 -15
ctg ctt cta atc acc cct tct ccc agc cct ctt cta ttt gat aga ggt     394
Leu Leu Leu Ile Thr Pro Ser Pro Ser Pro Leu Leu Phe Asp Arg Gly
    -10                  -5                 1                 5
ctg tcc ctc aga tca gca atg tct tagcccctct cctctcttcc attccttcct    448
Leu Ser Leu Arg Ser Ala Met Ser
                10
gttggtactc atttcttcta acttttaata aacatttagg tataatacat tacagtaagt    508
gctatttaga tacaaactta aaacatacta tatattttaa ggatctaaga atcctttara    568
rrrggcacat gactgaagta cctcagctgc gcagcctgta accagttttt ttaatgtaaa    628
agtaaraatg ccagccttaa cctabccctg carataaaag ctaacttta ttaataccag     688
ccctgaataa tggcactaat ccacactctt ccttaragtg atgctggaaa aataaaatca    748
ggggcttcag attaaaaaaa aaa                                            771
```

<210> 348
<211> 409
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 186..380

```
<221> sig_peptide
<222> 186..233
<223> Von Heijne matrix
      score 4
      seq FFLFLSFVLMYDG/LR

<221> polyA_signal
<222> 383..388

<221> polyA_site
<222> 396..409

<400> 348
ataaaagaag cagcaaatag aatttcccac aaagtaagtt gactctaaat cttaagtatt      60
acctagtttt ttaaaggttt gaatataata atgcagtatt tgcagtataa aaaggaagga     120
atttgtagag aatcattttg gtgctcaagt ctcttagcag tgccttattg cctcatagca     180
agaag atg ctg ggg ttt ttt ttg ttt ttg tcc ttt gta tta atg tat gat    230
      Met Leu Gly Phe Phe Leu Phe Leu Ser Phe Val Leu Met Tyr Asp
        -15                -10              -5
ggt ttg cgc ctt ttt ggc att ctt tca aca tgt cgt gta cat cac acc      278
Gly Leu Arg Leu Phe Gly Ile Leu Ser Thr Cys Arg Val His His Thr
    1              5               10               15
atg aat cag ttc cta att gat ata tct agc ttt acc tcc cga gtt aaa      326
Met Asn Gln Phe Leu Ile Asp Ile Ser Ser Phe Thr Ser Arg Val Lys
        20               25               30
aaa aaa atc ttt tta ttt tat gcc ttc awa ggt tgc ycg ttt car agt      374
Lys Lys Ile Phe Leu Phe Tyr Ala Phe Xaa Gly Cys Xaa Phe Gln Ser
        35               40               45
gcc aca taaataaaat gtttaacaaa aaaaaaaaa                               409
Ala Thr


<210> 349
<211> 613
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 69..458

<221> sig_peptide
<222> 69..233
<223> Von Heijne matrix
      score 4
      seq AALCGISLSQLFP/EP

<221> polyA_signal
<222> 564..569

<221> polyA_site
<222> 602..613

<400> 349
aagaacctga gcagcctgtc ttcagacaga gagaggccca cggctgtttc ttgaaaytgg      60
cgctggga atg gcc atg tgg aac agg cca tgb bag ang ctg cct cag cag     110
        Met Ala Met Trp Asn Arg Pro Xaa Xaa Xaa Leu Pro Gln Gln
        -55                -50              -45
cct cts sta gct gag ccc act gca gag ggg gag cca cac ctg ccc acg     158
Pro Leu Xaa Ala Glu Pro Thr Ala Glu Gly Glu Pro His Leu Pro Thr
```

```
        -40                 -35                 -30
ggc cgg gas byg act gag gcc aac cgc ttc gcc tat gct gcc ctc tgt      206
Gly Arg Xaa Xaa Thr Glu Ala Asn Arg Phe Ala Tyr Ala Ala Leu Cys
-25                 -20                 -15                 -10
ggc atc tcc ctg tcc cag tta ttt cct gaa ccc gaa cac agc tcc ttc      254
Gly Ile Ser Leu Ser Gln Leu Phe Pro Glu Pro Glu His Ser Ser Phe
                     -5                  1                   5
tgc aca gag ttc atg gca ggc ctg gtg ckm tgg ctg gag ttg tct gaa      302
Cys Thr Glu Phe Met Ala Gly Leu Val Xaa Trp Leu Glu Leu Ser Glu
            10                  15                  20
gct gtc ttg cca acc atg act gct ttt gcg agc ggc ctg gga ggt gaa      350
Ala Val Leu Pro Thr Met Thr Ala Phe Ala Ser Gly Leu Gly Gly Glu
    25                  30                  35
gga sca vma tgt gtt tgt tca aat ttt act gaa gga ccc cat ctt gaa      398
Gly Xaa Xaa Cys Val Cys Ser Asn Phe Thr Glu Gly Pro His Leu Glu
40                  45                  50                  55
gga cga ccc gac ggt gat cac tca gga cct tct gag ctt ctc act caa      446
Gly Arg Pro Asp Gly Asp His Ser Gly Pro Ser Glu Leu Leu Thr Gln
                60                  65                  70
gga tgg gca cta tgacscccgg gccagagtcc tcgtttgcca catgacctcc          498
Gly Trp Ala Leu
                75
ctgctccaag tgcccttgga ggagctggat gtccttgaaa agatgttcct ggagagcctg    558
aaggaaatca agaagagga atctgaaatg gccgaggcat cccraaaaaa aaaaa          613
```

<210> 350
<211> 986
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 12..638

<221> sig_peptide
<222> 12..263
<223> Von Heijne matrix
      score 4.19999980926514
      seq ITMLQMLALLGYG/LF

<221> polyA_signal
<222> 951..956

<221> polyA_site
<222> 975..985

<400> 350
```
accctatcaa g atg gtc aac ttc ccc cag aaa att gca ggt gaa ctc tat      50
            Met Val Asn Phe Pro Gln Lys Ile Ala Gly Glu Leu Tyr
                     -80                 -75
gga cct ctc atg ctg gtc ttc act ctg gtt gct atc cta ctc cat ggg      98
Gly Pro Leu Met Leu Val Phe Thr Leu Val Ala Ile Leu Leu His Gly
    -70                 -65                 60
atg aag acg tct gac act att atc cgg gag ggc acc ctg atg ggc aca      146
Met Lys Thr Ser Asp Thr Ile Ile Arg Glu Gly Thr Leu Met Gly Thr
-55                 -50                 -45                 -40
gcc att ggc acc tgc ttc ggc tac tgg ctg gga gtc tca tcc ttc att      194
Ala Ile Gly Thr Cys Phe Gly Tyr Trp Leu Gly Val Ser Ser Phe Ile
            -35                 -30                 -25
tac ttc ctt gcc tac ctg tgc aac gcc cag atc acc atg ctg cag atg      242
```

```
Tyr Phe Leu Ala Tyr Leu Cys Asn Ala Gln Ile Thr Met Leu Gln Met
        -20              -15              -10
ttg gca ctg ctg ggc tat ggc ctc ttt ggg cat tgc att gtc ctg ttc      290
Leu Ala Leu Leu Gly Tyr Gly Leu Phe Gly His Cys Ile Val Leu Phe
        -5              1               5
atc acc tat aat atc cac ctc cgc gcc ctc ttc tac ctc ttc tgg ctg      338
Ile Thr Tyr Asn Ile His Leu Arg Ala Leu Phe Tyr Leu Phe Trp Leu
10              15              20              25
ttg gtg ggt gga ctg tcc aca ctg cgc atg gta gca gtg ttg gtg tct      386
Leu Val Gly Gly Leu Ser Thr Leu Arg Met Val Ala Val Leu Val Ser
        30              35              40
cgg acc gtg ggc ccc aca cad cgg mtg ctc ctc tgt ggc acc ctg gct      434
Arg Thr Val Gly Pro Thr Xaa Arg Xaa Leu Leu Cys Gly Thr Leu Ala
        45              50              55
gcc cta cac atg ctc ttc ctg ctc tat ctg cat ttt gcc tac cac aaa      482
Ala Leu His Met Leu Phe Leu Leu Tyr Leu His Phe Ala Tyr His Lys
        60              65              70
dtg gta dag ggg atc ctg gac aca ctg gag ggc ccc aac atc ccg ccc      530
Xaa Val Xaa Gly Ile Leu Asp Thr Leu Glu Gly Pro Asn Ile Pro Pro
        75              80              85
atc cag agg gtc ccc aga gac atc cct gcc atg ctc cct gct gct cgg      578
Ile Gln Arg Val Pro Arg Asp Ile Pro Ala Met Leu Pro Ala Ala Arg
90              95              100             105
ctt ccc acc acc gtc ctc aac gcc aca gcc aaa gct gtt gcg gtg acc      626
Leu Pro Thr Thr Val Leu Asn Ala Thr Ala Lys Ala Val Ala Val Thr
        110             115             120
ctg cag tca cac tgaccccacc tgaaattctt ggccagtcct ctttcccgca      678
Leu Gln Ser His
        125
gctgcagaga ggargaasac tattaaagga cagtcctgat gacatgtttc gtagatgggg      738
tttgcagctg ccactgagct gtagctgcgt aagtacctcc ttgatgcctg tcggcacttc      798
tgaaaggcac aaggccaaga actcctggcc aggactgcaa ggctctgcag ccaatgcaga      858
aaatgggtca gctcctttga gaacccctcc ccacctaccc cttccttcct ctttatctct      918
cccacattgt cttgctaaat atagacttgg taattaaaat gttgattgaa gtctggaaaa      978
aaaaaaat                                                              986
```

```
<210> 351
<211> 1447
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 282..389

<221> sig_peptide
<222> 282..332
<223> Von Heijne matrix
      score 3.5
      seq RWWCFHLQAEASA/HP

<221> polyA_signal
<222> 1413..1418

<221> polyA_site
<222> 1437..1447

<400> 351
ataataatat ctaaaaagct aaattttaaa taccagcttt acataaatga ttgtkgactc      60
tggtctgtkt ctgacacctt tccagaaaaa agtcaattgt tcaggtacac caaagaggaa      120
```

```
gaagagctgt ggaggccacc ctctacaaag ctttatagaa cttctggatc taactcacaa    180
acaagcttcc agaagagact agagacctta ggccaggaga tgaaggagtt cagtagcaaa    240
gtcacacctg tccaattccc tgagctttgc tcactcagct a atg gga tgg caa agg    296
                                             Met Gly Trp Gln Arg
                                                          -15
tgg tgg tgc ttt cat ctt cag gca gaa gcc tct gcc cat ccc cct caa    344
Trp Trp Cys Phe His Leu Gln Ala Glu Ala Ser Ala His Pro Pro Gln
        -10                 -5                  1
ggg ctg cag gcc caa ttc tca tgc tgc cct tgg gtg ggc atc tgt        389
Gly Leu Gln Ala Gln Phe Ser Cys Cys Pro Trp Val Gly Ile Cys
5                  10                  15
taacaaadga aaacgtctgg gtggcggcag casctttgct ctgagtgcct acaaagctaa    449
tgcttggtgc tagaaacatc atcattatta aacttcagaa aagcagcagc catgttcagt    509
caggctcatg ctgcctcact gcttaagtgc ctgcaggagc cgcctgccaa rctcccccttc    569
ctacacctgg cacactgggg tctgcacaag gctttgtcaa ccaaaracag cttcccccww    629
ttgattgcct gtagactttg gagccaaraa acactcctgt tgactctaca cacacttcag    689
gtggtttgtg cttcaaagtc attgatgcaa cttgaaagga aacagtttaa tggtggaaat    749
gaactaccat ttataacttc tgttttttta ttgagaaaat gattcacgaa kkcca?atca    809
gattgccagg aagaaatagg acgtgacggt actgggccct gtgattctcc cagccuttgc    869
agtccgctag gtgagaggaa aagctcttta cttccgcccc tggcagggac ttctgggtta    929
tgggagaaac cagagatggg aatgaggaaa atatgaacta cagcagaagc ccctgggcag    989
ctgtgatgga gcccctgaca ttactcttct tgcatctgtc ctgccttctt tccctctgcg   1049
aggcagtggg gtgggattca gagtgcttag tctgctcact gggagaagaa gagttcctgc   1109
gcatgcaagc cctgctgtgt ggctgtcgtt tacatttggg aggtgtcctg tatgtctgta   1169
cgttggggac tgcctgtatt tggaagattt aaaaacctag catcctgttc tcaccctcta   1229
agctgcattg agaaatgact cgtctctgta tttgtattaa gccttaacac ttttcttaag   1289
tgcattcggt gccaacattt tttagagctg taccaaaaca aaaagcctgt actcacatca   1349
camtgtcatt ttgataggag cgttttgtta tttttacaag gcagaatggg gtgtaacagt   1409
tgaattaaac ttagcaatca cgtgctcaaa aaaaaaaa                           1447
```

<210> 352
<211> 1641
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 208..339

<221> sig_peptide
<222> 208..294
<223> Von Heijne matrix
        score 5.59999990463257
        seq LFLQLLVSHEIVC/AT

<221> polyA_site
<222> 1631..1641

<400> 352
```
agaaccgtga tgggaagatg gacaaggaag agaccaaaga ctggatcctt ccctcagact     60
atgatcgatgc agaggcagaa gccaggcacc tggtctatga atcagaccaa aacaaggatg    120
gcaagcttac caaggaggag atcgttgaca gtatgacttt atttgttggc agccaggcca    180
cagattttgg ggaggcctta gtacggc atg atg agt tct gag cta cgg agg aac    234
                                 Met Met Ser Ser Glu Leu Arg Arg Asn
                                                              -25
cct cat ttc ctc aaa agt aat tta ttt tta cag ctt ctg gtt tca cat    282
Pro His Phe Leu Lys Ser Asn Leu Phe Leu Gln Leu Leu Val Ser His
-20                 -15                 -10                 -5
gaa att gtt tgc gct act gag act gtt act aca aac ttt tta aga cat    330
Glu Ile Val Cys Ala Thr Glu Thr Val Thr Thr Asn Phe Leu Arg His
```

```
            1              5               10
gaa aag gcg taatgaaaac catcccgtcc ccattcctcc tcctctctga        379
Glu Lys Ala
         15
gggactggag ggaagccgtg cttctgagga acaactctaa ttagtacact tgtgtttgta  439
ratttacacw wtgtattatg tattaacatg gcgtgtttat ttttgtattt ttctctggtt  499
gggagtatka tatgaaggat caaratcctc aactcacaca tgtaracaaa cattasctct  559
ttactctttc tcaacccctt wtatgatttt aataattctc acttaactaa ttttgtaagc  619
ctgagatcaa taagaaatgt tcaggagaga ggaaagaaaa aaaatatatg ctccacaatt  679
tatatttaga gagagaacac ttagtcttgc ctgtcaaaaa gtccaacatt tcataggtag  739
taggggccac atattacatt cagttgctat aggtccagca actgaacctg ccattacctg  799
ggcaaggaaa gatccctttg ctctaggaaa gcttggccca aattgatttt cttctttttc  859
cccctgtagg actgactgtt ggctaatttt gtcaagcaca gctgtggtgg aagagttag   919
ggccagtgtc ttgaaaatca atcaagtagt gaatgtgatc tctttgcara gctatagata  979
gaaacagctg gaaaactaaa ggaaaaatac aagtgttttc ggggcataca tttttttttct 1039
gggtgtgcat ctgttgaaat gctcaagact taattatttg cctttttgaaa tcactgtaaa 1099
tgcccccatc cggttcctct tcttcccarg tgtgccaagg aattaatctt ggtttcacta 1159
caattaaaat tcactccttt ccaatcatgt cattgaaagt gcctttaacg aaagaaatgg 1219
tcactgaatg gguattctct taagaaaccc tgagattaaa aaaagactat ttggataact 1279
tataggaaag cctagaacct cccagtagag tggggatttt tttcttcttc cctttctctt 1339
ttggacaata gttaaattag cagtattagt tatgagtttg gttgcagtgt tcttatcttg 1399
tgggctgatt tccaaaaacc acatgctgct gaatttacca gggatcctca tacctcacaa 1459
tgcaaaccac ttactaccag gccttttttct gtgtccactg gagagcttga gctcacactc 1519
aaagatcaga ggacctacag agagggctct ttggtttgag gaccatggct tacctttcct 1579
gcctttgacc catcacaccc catttcctcc tctttccctc tccccgctgc caaaaaaaaa 1639
aa                                                             1641
```

<210> 353
<211> 884
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 69..557

<221> sig_peptide
<222> 69..224
<223> Von Heijne matrix
      score 4.69999980926514
      seq LGLALGRLEGGSA/RH

<221> polyA_signal
<222> 849..854

<221> polyA_site
<222> 870..883

<400> 353
```
attggctccg gatcgtgcgt gaggcggctt cgtgggcagc gagagtcaca gacaagacag    60
caagcagg atg gag cac tac cgg aaa gct ggc tct gta gag ctc cca gcg    110
         Met Glu His Tyr Arg Lys Ala Gly Ser Val Glu Leu Pro Ala
            -50              -45                  -40
cct tcc cca atg ccc cag cta cct cct gat acc ctt gag atg cgg gtc    158
Pro Ser Pro Met Pro Gln Leu Pro Pro Asp Thr Leu Glu Met Arg Val
      -35              -30              -25
cga gat ggc agc aaa att cgc aac ctg ctg ggg ttg gct ctg ggt cgg    206
Arg Asp Gly Ser Lys Ile Arg Asn Leu Leu Gly Leu Ala Leu Gly Arg
      -20              -15              -10
ttg gag ggc ggc agt gct cgg cat gta gtg ttc cca ggt tct ggc agg    254
```

```
          Leu Glu Gly Gly Ser Ala Arg His Val Val Phe Ser Gly Ser Gly Arg
               -5                   1                   5                  10
          gct gca gga aag gct gtc agc tgc gct gag att gtc aag cgg cgg gtc      302
          Ala Ala Gly Lys Ala Val Ser Cys Ala Glu Ile Val Lys Arg Arg Val
                            15                  20                  25
          ccg ggc ctg cac cag ctc acc aag cta ckt ttc ctt caa act gag gac      350
          Pro Gly Leu His Gln Leu Thr Lys Leu Xaa Phe Leu Gln Thr Glu Asp
                        30                  35                  40
          agc tgg gtc cca scc tca cct gac aca ggg cta rac ccc ctc aca gtg      398
          Ser Trp Val Pro Xaa Ser Pro Asp Thr Gly Leu Xaa Pro Leu Thr Val
                    45                  50                  55
          cgc cgc cat gtg cct gca ktg tgg gtg ctg ctc asc cgg gac ccc ctg      446
          Arg Arg His Val Pro Ala Xaa Trp Val Leu Leu Xaa Arg Asp Pro Leu
                60                  65                  70
          gac ccc aat gag tgt ggt tac caa ccc cca gga gca ccc cct ggc ctg      494
          Asp Pro Asn Glu Cys Gly Tyr Gln Pro Pro Gly Ala Pro Pro Gly Leu
          75                  80                  85                  90
          ggt tcc atg ccc agc tcc agc tgt ggc cct cgt tcc cra aaa agg gct      542
          Gly Ser Met Pro Ser Ser Ser Cys Gly Pro Arg Ser Xaa Lys Arg Ala
                        95                  100                 105
          cra rac acc cga tcg tgaaaacctg ctgasccagc ctgttctccg ggcctraatg     597
          Xaa Xaa Thr Arg Ser
                        110
          tctggggtgc ttgtgccttt tctranaagc gttgtgaskg ctcaacatcc ccatcaaggt     657
          ttgagtccac aaaagtggac ctccctatca tgcttcccct tccctctagc atgtgggaag     717
          ggactgctgt gaagaatgac agatgtgggg cctctgccaa gttctgcatt gctaaataag     777
          ggcttcctct gccttctacc tacagtgcat ttgaactgcc ttctgaaaga ggtccakgga     837
          gggatttagg aaataaagtt tctacctatt tgaaaaaaaa aaaacac                   884
```

<210> 354
<211> 729
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 134..325

<221> sig_peptide
<222> 134..274
<223> Von Heijne matrix
      score 5.90000009536743
      seq TWLGLLSFQNLHC/FP

<221> polyA_site
<222> 718..729

<400> 354
```
          atcattttct tatccctgct gatttcaaac cttcccatgg tttagaagca taacctgtaa      60
          tgtaatgcaa gtcccctaac tccctggttg ctaacattaa cttccttaag taataatcaa     120
          tgaaagavat tct atg cat ggt ttt gaa ata ata tcc ttg aaa gag gaa        169
                      Met His Gly Phe Glu Ile Ile Ser Leu Lys Glu Glu
                            -45                 -40
          tca cca tta gga aag gtg agt cag ggt cct ttg ttt aat gtg act agt      217
          Ser Pro Leu Gly Lys Val Ser Gln Gly Pro Leu Phe Asn Val Thr Ser
          -35                 -30                 -25                 -20
          ggc tca tca tca cca gtg acc tgg ttg ggc cta ctc tcc ttc cag aac      265
          Gly Ser Ser Ser Pro Val Thr Trp Leu Gly Leu Leu Ser Phe Gln Asn
                        -15                 -10                 -5
          ctg cat tgc ttc cca gac ctc ccc act gag atg cct cta ara gcc aaa      313
```

```
Leu His Cys Phe Pro Asp Leu Pro Thr Glu Met Pro Leu Xaa Ala Lys
            1               5                  10
gga ktc aac act tgagcctagg gtgggctaca acaaaaratt ctaatttacc·         365
Gly Xaa Asn Thr
            15
ttgcttcatc taggtccagg ccccaaktag cttgctgaag gaacttaaaa agtagctgtt    425
atttattgta ttgtataasc taaaaacatt tattttgtt gaatcraaac aattccatgt     485
ascaatcttt tttctgttca cggtgtttgt gataaaacct taaattccgc aagcatcagt    545
tttttgaaaa aatgggaatt gaccggatag wwacaggcaa agwtataaat agctacaaca    605
tcatttaact tttataaaca tgccttctct ctattgaara catctgatat ttttgctgga    665
aagttggatc tatcctcagt aactctgcca tgaattcctg tttcckggtt ccaaaaaaaa    725
aaaa                                                                729
```

```
<210> 355
<211> 1013
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 78..731

<221> sig_peptide
<222> 78..227
<223> Von Heijne matrix
      score 5.09999990463257
      seq RTALILAVCCGSA/SI

<221> polyA_site
<222> 1002..1013

<400> 355
agtttccaag ggaaggagca gcgtgtggga aagcacagaa gagtgagaag gaagcgacta     60
aatttattt actttct atg cat cat ggc ctc aca cca ctg tta ctt ggt        110
                    Met His His Gly Leu Thr Pro Leu Leu Leu Gly
                    -50             -45                     -40
gta cat gag caa aaa cag caa gtg gtg aaa ttt tta atc aag aaa aaa      158
Val His Glu Gln Lys Gln Gln Val Val Lys Phe Leu Ile Lys Lys Lys
        -35                 -30                     -25
gca aat tta aat gca ctg gat aga tat gga aga act gct ctc ata ctt     206
Ala Asn Leu Asn Ala Leu Asp Arg Tyr Gly Arg Thr Ala Leu Ile Leu
        -20                 -15                     -10
gct gta tgt tgt gga tcg gca agt ata gtc agc ctt cta ctt gag caa     254
Ala Val Cys Cys Gly Ser Ala Ser Ile Val Ser Leu Leu Leu Glu Gln
        -5                  1                   5
aac att gat gta tct tct caa gat cta tct gga cag acg gcc aaa aag     302
Asn Ile Asp Val Ser Ser Gln Asp Leu Ser Gly Gln Thr Ala Lys Lys
10                  15                  20                  25
tat gct gtt tct agt cgt cat aat gta att tgc cag tta ctt tct gac     350
Tyr Ala Val Ser Ser Arg His Asn Val Ile Cys Gln Leu Leu Ser Asp
                30                  35                  40
tac aaa raa aaa cag atr cta aaa gtc tct tct gaa aac agc aat cca     398
Tyr Lys Xaa Lys Gln Xaa Leu Lys Val Ser Ser Glu Asn Ser Asn Pro
                45                  50                  55
raa caa gac tta aag ctg aca tca gag gaa gag tca caa agg ctt aaa     446
Xaa Gln Asp Leu Lys Leu Thr Ser Glu Glu Glu Ser Gln Arg Leu Lys
        60                  65                  70
gga agt gaa aat agc cag cca gag gaa atg tct caa gaa cca gaa ata     494
Gly Ser Glu Asn Ser Gln Pro Glu Glu Met Ser Gln Glu Pro Glu Ile
        75                  80                  85
```

```
aat arg ggt ggt gat aga aag gtt gaa raa raa atg aar aag cac gga      542
Asn Xaa Gly Gly Asp Arg Lys Val Glu Xaa Xaa Met Lys Lys His Gly
90                  95                  100                 105
agt wct cat atg gga ttc cca raa aac ctg mct aac ggt gcc act gct      590
Ser Xaa His Met Gly Phe Pro Xaa Asn Leu Xaa Asn Gly Ala Thr Ala
                110                 115                 120
gac aat ggt gat gat gga tta att ccm cca rgg aaa asc ara aca cct      638
Asp Asn Gly Asp Asp Gly Leu Ile Pro Pro Xaa Lys Xaa Xaa Thr Pro
            125                 130                 135
gaa agc cas caa ttt cct gac act gag aat gaa cag tat cac agg gac      686
Glu Ser Xaa Gln Phe Pro Asp Thr Glu Asn Glu Gln Tyr His Arg Asp
        140                 145                 150
ttt tct ggc cat ccc mac ttt ccc acd acc ctt ccc atc aaa cag         731
Phe Ser Gly His Pro Xaa Phe Pro Thr Thr Leu Pro Ile Lys Gln
    155                 160                 165
tgatgaacaa aatgatactc hsaagcmmct ttctgaagam caaacactg gaatattaca    791
agatgagatt ctgattcatg aagaaaagca gatagaagtg gctgaaaatg aattctgagc    851
tttctcttag ttataaraaa gaaaaagacc tcttgcatga aaatagtacg ttgcaggaag    911
aaattgtcat gctaaractg gaactagack taatgaaaca tcagagccag ctaararaaa    971
araaatattt ggaggaaatt gaaagtgtgg aaaaaaaaaa aa                      1013
```


<210> 356
<211> 973
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 46..693

<221> sig_peptide
<222> 46..90
<223> Von Heijne matrix
      score 7.59999990463257
      seq CVLVLAAAAGAVA/VF

<221> polyA_signal
<222> 937..942

<221> polyA_site
<222> 962..973

<400> 356
```
aagcggctgg tccccggaag ttggacgcat gcgccgtttc tctgc atg gtg tgc gtt     57
                                                 Met Val Cys Val
                                                     -15
ctc gtt cta gct gcg gcc gca gga gct gtg gcg gtt ttc cta atc ctg     105
Leu Val Leu Ala Ala Ala Ala Gly Ala Val Ala Val Phe Leu Ile Leu
    -10                 -5                  1               5
cga ata tgg gta gtg ctt cgt tcc atg gac gtt acg ccc cgg gag tct     153
Arg Ile Trp Val Val Leu Arg Ser Met Asp Val Thr Pro Arg Glu Ser
                10                  15                  20
ctc agt atc ttg gta gtg gct ggg tcc ggt ggg cat acc act gag atc     201
Leu Ser Ile Leu Val Val Ala Gly Ser Gly Gly His Thr Thr Glu Ile
        25                  30                  35
ctg agg ctg ctt ggg agc ttg tcc aat gcc tac tca cct aga cat tat     249
Leu Arg Leu Leu Gly Ser Leu Ser Asn Ala Tyr Ser Pro Arg His Tyr
    40                  45                  50
gtc att gct gac act gat gaa atg agt gcc aat aaa ata aat tct ttt     297
Val Ile Ala Asp Thr Asp Glu Met Ser Ala Asn Lys Ile Asn Ser Phe
```

```
           55                      60                      65
gaa cta rat cga gsk gat aga rac cct agt aac atg twt acc aaa tac       345
Glu Leu Xaa Arg Xaa Asp Arg Xaa Pro Ser Asn Met Xaa Thr Lys Tyr
70                      75                      80                      85
tac att cac cga att cca ara agc cgg gag gtt cag cag tcc tgg ccc       393
Tyr Ile His Arg Ile Pro Xaa Ser Arg Glu Val Gln Gln Ser Trp Pro
                        90                      95                      100
tcc acc gtt tyc acc acc ttg cac tcc atg tgg ctc tcc ttk ccc cta       441
Ser Thr Val Xaa Thr Thr Leu His Ser Met Trp Leu Ser Xaa Pro Leu
                   105                      110                      115
att cac agg gtg aag cca rat ttg gtg ttg tgt aac gga cca gga aca       489
Ile His Arg Val Lys Pro Xaa Leu Val Leu Cys Asn Gly Pro Gly Thr
              120                      125                      130
tgt gty cct atc tgt gta tct gcc ctt ctc ctt ggg ata cta gga ata       537
Cys Val Pro Ile Cys Val Ser Ala Leu Leu Leu Gly Ile Leu Gly Ile
         135                      140                      145
aag aaa gtg atc att gtc tac gtt gaa agc atc tgc cgt gta aaa acs       585
Lys Lys Val Ile Ile Val Tyr Val Glu Ser Ile Cys Arg Val Lys Thr
150                      155                      160                      165
tta tcc atg tcc gga aag att ctg ttt cat ctc tca aat tac ttc att       633
Leu Ser Met Ser Gly Lys Ile Leu Phe His Leu Ser Asn Tyr Phe Ile
                   170                      175                      180
gtt cag tgg ccg gct ctg aaa gaa aag tat ccc aaa tcg gtg tac ctt       681
Val Gln Trp Pro Ala Leu Lys Glu Lys Tyr Pro Lys Ser Val Tyr Leu
              185                      190                      195
ggg cga att gtt tgacaaatgg caactgactt ctttagaatt ttgcasttaa          733
Gly Arg Ile Val
         200
cagtartatg tactcaaatt ggggggaaaa aaaccctaca tgtttcttgt aaaggcgtct     793
gacagtcctg araattattg atggtaagga ataaaaaatg twcagatrac tcagtgaara     853
aactgaggct tctcttatga aacaaacatt gataaacgta actacyaaat gtttatgcct     913
ctgtaaacca aatttctttt ctarataaaa atatgtatta ctacctgcaa aaaaaaaaaa     973
```

<210> 357
<211> 868
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 126..527

<221> sig_peptide
<222> 126..182
<223> Von Heijne matrix
      score 3.90000009536743
      seq ILFHGVFYAGGFA/IV

<221> polyA_signal
<222> 834..839

<221> polyA_site
<222> 856..867

<400> 357
```
actggaagaa ctcgtcatgc tctttgtagc gtggtgctcc tgttgctcac aggacaactt     60
gcctttgatg attttcaaga gagttgtgct atgatgtggc aaagtatgca ggaagcaggc     120
ggtca atg cct ctg gga gca agg atc ctt ttc cac ggt gtg ttc tat gcc    170
      Met Pro Leu Gly Ala Arg Ile Leu Phe His Gly Val Phe Tyr Ala
          -15                     -10                      -5
```

```
ggg ggc ttt gcc att gtg tat tac ctc att caa aag ttt cat tcc agg      218
Gly Gly Phe Ala Ile Val Tyr Tyr Leu Ile Gln Lys Phe His Ser Arg
                1                   5                  10
act tta tat tac aag ttg gca gtg gar cag ctg car arc cat ccc gag      266
Thr Leu Tyr Tyr Lys Leu Ala Val Glu Gln Leu Gln Xaa His Pro Glu
        15                  20                  25
gca cag gaa gct ctg ggc cct cct ctc aac atc cat tat ctc aag ctc      314
Ala Gln Glu Ala Leu Gly Pro Pro Leu Asn Ile His Tyr Leu Lys Leu
        30                  35                  40
atc gac agg gaa aac ttc gtg gac att gtt rat gcc aag ttg aaa att      362
Ile Asp Arg Glu Asn Phe Val Asp Ile Val Xaa Ala Lys Leu Lys Ile
45                  50                  55                  60
cct gtc tct gga tcc aaa tca gag ggc ctt ctc tac gtc cac tca tcc      410
Pro Val Ser Gly Ser Lys Ser Glu Gly Leu Leu Tyr Val His Ser Ser
                65                  70                  75
aga ggt ggc ccc ttt cag agg tgg cac ctt gac gag gtc ttt tta gag      458
Arg Gly Gly Pro Phe Gln Arg Trp His Leu Asp Glu Val Phe Leu Glu
                80                  85                  90
ctc aag gat ggt cag cag att cct gtg ttc aag ctc agt ggg gaa aac      506
Leu Lys Asp Gly Gln Gln Ile Pro Val Phe Lys Leu Ser Gly Glu Asn
        95                  100                 105
ggt gat gaa gtg aaa aag gag tagagacgac ccagaagacc cagcttgctt         557
Gly Asp Glu Val Lys Lys Glu
        110                 115
ctagtccatc cttccctcat ctctaccata tggccactgg ggtggtggcc catctcagtg    617
acagacactc ctgcaaccca gktttccagc caccagtggg atgatggtat gtgccagcac    677
atggtaattt tggtgtaatt ctaacttggg cacaacgaat gctattgtc attttaaac     737
tgaatccgaa agaaactcct attataaatt taagataatg taatgtattt gaaagtgctt    797
tgtataaaaa agcacatgat aaaaggaatc agaattaata aaatgtttgt tgatctttaa    857
aaaaaaaaaa h                                                         868


<210> 358
<211> 519
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 66..320

<221> sig_peptide
<222> 66..113
<223> Von Heijne matrix
      score 3.5
      seq TALAAXTWLGVWG/VR

<221> polyA_signal
<222> 490..495

<221> polyA_site
<222> 508..519

<400> 358
aattagcgcg taacgcasag actgcttgct gcggcagaca cgccagakgt gcagctccag      60
cagca atg gca gtg acg gcg ttg gcg gcg mrg acg tgg ctt ggc gtg tgg   110
      Met Ala Val Thr Ala Leu Ala Ala Xaa Thr Trp Leu Gly Val Trp
           -15                 -10                 -5
ggc gtg agg acc atg caa gcc cga ggc ttc ggc tcg gat cag tcc gag   158
Gly Val Arg Thr Met Gln Ala Arg Gly Phe Gly Ser Asp Gln Ser Glu
         1                   5                  10                  15
```

```
aat gtc gac cgg ggc gcg ggc tcc atc cgg gaa gcc ggt ggg gcc ttc      206
Asn Val Asp Arg Gly Ala Gly Ser Ile Arg Glu Ala Gly Gly Ala Phe
            20              25              30
gga aag aga gag cag gct gaa gag gaa cga tat ttc cga gca cag agt      254
Gly Lys Arg Glu Gln Ala Glu Glu Glu Arg Tyr Phe Arg Ala Gln Ser
        35              40              45
aca gaa caa ctg gca rct ttg aaa aaa crc cat gaa gaa gar atc gtt      302
Thr Glu Gln Leu Ala Xaa Leu Lys Lys Xaa His Glu Glu Glu Ile Val
        50              55              60
cat cat aga gaa gga gat tgagcgtctg cagaaagaaa ttgagcgcca            350
His His Arg Glu Gly Asp
        65
taagcagaag atcaaaatgc tagaacatga tgattaagtg cacaccgtgt gccatagaat    410
ggcacatgtc attgcccact tctgtgtaaa catggttctg gtttaactaa tatttgtctg    470
tgtgctacta acagattata ataaattgtc atcagtgaaa aaaaaaaaa                519
```

```
<210> 359
<211> 1028
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 73..948

<221> sig_peptide
<222> 73..159
<223> Von Heijne matrix
      score 4.40000009536743
      seq IVLHLVLQGMVYT/EY

<221> polyA_site
<222> 1016..1028

<400> 359
agctttaaag gcctggccag gggaggagca cagatatttt cctgtataat tccagaatgt    60
cttcagagag cc atg cat gga ttg ctt cat tac ctt ttc cat acg aga aac    111
            Met His Gly Leu Leu His Tyr Leu Phe His Thr Arg Asn
                           -25                   -20
cac acc ttc att gtc ctg cac ctg gtc ttg caa ggg atg gtt tat act      159
His Thr Phe Ile Val Leu His Leu Val Leu Gln Gly Met Val Tyr Thr
        -15                 -10                 -5
gag tac acc tgg gaa gta ttt ggc tac tgt cag gag ctg gag ttg tcc      207
Glu Tyr Thr Trp Glu Val Phe Gly Tyr Cys Gln Glu Leu Glu Leu Ser
1           5               10              15
ttg cat tac ctt ctt ctg ccc tat ctg ctg cta ggt gta aac ctg ttt      255
Leu His Tyr Leu Leu Leu Pro Tyr Leu Leu Leu Gly Val Asn Leu Phe
            20              25              30
ttt ttc acc ctg act tgt gga acc aat cct ggc att ata aca aaa gca      303
Phe Phe Thr Leu Thr Cys Gly Thr Asn Pro Gly Ile Ile Thr Lys Ala
        35              40              45
aat gaa tta tta ttt ctt cat gtt tat gaa ttt gat gaa ktg atg ttt      351
Asn Glu Leu Leu Phe Leu His Val Tyr Glu Phe Asp Glu Xaa Met Phe
        50              55              60
cca aaa aac gtg agg tgc tct act tgt gat tta agg aaa cca gct cga      399
Pro Lys Asn Val Arg Cys Ser Thr Cys Asp Leu Arg Lys Pro Ala Arg
65              70              75              80
tcc aas cac tgc akt gtg tgt aac tgg tgt gtg cac cgt ttc rac cat      447
Ser Xaa His Cys Xaa Val Cys Asn Trp Cys Val His Arg Phe Xaa His
            85              90              95
```

```
cac tgt gtt tgg gtg aac aac tgc atc ggg gcc tgg aac atc agg tmc      495
His Cys Val Trp Val Asn Asn Cys Ile Gly Ala Trp Asn Ile Arg Xaa
            100                     105                 110
ttc ctc atc tac gtc ttg acc ttg acg gcc tcg gct gcc acc gtc gcc      543
Phe Leu Ile Tyr Val Leu Thr Leu Thr Ala Ser Ala Ala Thr Val Ala
            115                     120                 125
att gtg agc acc act ttt ctg gtc cac ttg gtg gtg atg tca gat tta      591
Ile Val Ser Thr Thr Phe Leu Val His Leu Val Val Met Ser Asp Leu
            130                     135                 140
tac cag gag act tac atc gat gac ctt gga cac ctc cat gtt atg gac      639
Tyr Gln Glu Thr Tyr Ile Asp Asp Leu Gly His Leu His Val Met Asp
145                     150                 155                 160
acg gtc ttt ctt att cag tac ctg ttc ctg act ttt cca cgg att gtc      687
Thr Val Phe Leu Ile Gln Tyr Leu Phe Leu Thr Phe Pro Arg Ile Val
                    165                     170                 175
ttc atg ctg ggc ttt gtc gtg gtt ctg arc ttc ctc ctg ggt ggc tac      735
Phe Met Leu Gly Phe Val Val Val Leu Xaa Phe Leu Leu Gly Gly Tyr
                180                     185                 190
ctg ttg ttt gtc ctg tat ctg gcg gcc acc aac cag act act aac gag      783
Leu Leu Phe Val Leu Tyr Leu Ala Ala Thr Asn Gln Thr Thr Asn Glu
            195                     200                 205
tgg tac aga rgt gac tgg gcc tgg tgc cag cgt tgt ccc ctt gtg gcc      831
Trp Tyr Arg Xaa Asp Trp Ala Trp Cys Gln Arg Cys Pro Leu Val Ala
            210                     215                 220
tgg cct ccg tca gca gar ccc caa gtc cac cgg aac att cac tcc cat      879
Trp Pro Pro Ser Ala Glu Pro Gln Val His Arg Asn Ile His Ser His
225                     230                 235                 240
ggg ctt cgg arc aac ctt caa gar atc ttt cta cct gcc ttt cca tgt      927
Gly Leu Arg Xaa Asn Leu Gln Glu Ile Phe Leu Pro Ala Phe Pro Cys
                245                     250                 255
cat gag agg aag aaa caa gaa tgacmagtgt atgactgcct ttgagctgta      978
His Glu Arg Lys Lys Gln Glu
                260
gttcccgttt atttacacat gtggatcctc gtttttccaaa aaaaaaaaaa      1028
```

```
<210> 360
<211> 452
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 69..434

<221> sig_peptide
<222> 69..236
<223> Von Heijne matrix
      score 4.90000009536743
      seq FACVPGASPTTLA/FP

<221> polyA_signal
<222> 419..424

<221> polyA_site
<222> 441..452

<400> 360
acagcgtgas tcgcccgcca gaagaatatg aaaaagcaga gcganctcgg ttaagggaaa      60
gcgccgag atg acg ggc ttt ctg ctg ccg ccc gca agc aga ggg act cgg      110
         Met Thr Gly Phe Leu Leu Pro Pro Ala Ser Arg Gly Thr Arg
```

```
                 -55                    -50                    -45
aga tca tgc agc aga agc aga aaa agg caa acg aga aga agg agg aac        158
Arg Ser Cys Ser Arg Ser Arg Lys Arg Gln Thr Arg Arg Arg Arg Asn
        -40                    -35                    -30
cca agt agc ttt gtg gct tcg tgt cca acc ctc ttg ccc ttc gcc tgt       206
Pro Ser Ser Phe Val Ala Ser Cys Pro Thr Leu Leu Pro Phe Ala Cys
        -25                    -20                    -15
gtg cct gga gcc agt ccc acc acg ctc gcg ttt cct cct gta ktg ctc       254
Val Pro Gly Ala Ser Pro Thr Thr Leu Ala Phe Pro Pro Val Xaa Leu
-10                     -5                      1                      5
aca ggt ccc avc acc gat ggc att ccc ttt gcc ctr nak tct gca gcg       302
Thr Gly Pro Xaa Thr Asp Gly Ile Pro Phe Ala Leu Xaa Ser Ala Ala
                10                     15                     20
ggt ccc ttt tgt gct tcc ttc ccc tca ggt avc ctc tct ccc cct ggg       350
Gly Pro Phe Cys Ala Ser Phe Pro Ser Gly Xaa Leu Ser Pro Pro Gly
        25                     30                     35
cca ctc ccg ggg gtg agg ggg tta ccc ctt ccc agt gtt ttt tat tcc       398
Pro Leu Pro Gly Val Arg Gly Leu Pro Leu Pro Ser Val Phe Tyr Ser
        40                     45                     50
tgt ggg gct cac ccc aaa gta tta aaa gta gct ttg taattcaaaa            444
Cys Gly Ala His Pro Lys Val Leu Lys Val Ala Leu
55                      60                     65
aaaaaaaa                                                              452
```

<210> 361
<211> 875
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 628..804

<221> sig_peptide
<222> 628..711
<223> Von Heijne matrix
      score 4.19999980926514
      seq LMPVIPALQEAXA/GG

<221> polyA_site
<222> 864..875

<400> 361

```
aaagatggac accgcggagg aagacatatg tagagtgtgt cggtcagaag gaacacctga      60
gaaaccgctt tatcatcctt gtgtatgtac tggcagtatt aagttngtcc atcaagaatg     120
cttagttcaa tggctgaaac acagtcgaaa agaatactgt gaattatgca agcacagatt     180
tgcttttaca ccaatttatt ctccagatat gccttcacgg cttccaattc aagacatatt     240
tgctggactg gttacaagta ttggcactgc aatacgatat tggtttcatt atacacttgt     300
ggcctttgca tggttgggag ttgttcctct tacagcatgt gagtattcat gcctctgatt     360
ggagttattt aaacattgca taactactta atattataaa gcaatattgc atcatattat     420
tatttgactg atgtttagtt atttgatgtc agagtgtcat gtattaggaa agccttactt     480
araaratgtt catcggaact aaraatgakt ttaacaggtc agttttttga gtgaatgtgg     540
gaaaraacac agcatacaga atggctaacc atgaaagttc atgaaagcgt kgaaaaaatc     600
aaatcaaatc ataattagat atgaagt atg cta rag ctt tca agg gct aca aaa     654
                                Met Leu Xaa Leu Ser Arg Ala Thr Lys
                                         -25                    -20
rac ggc cgg gcg cgg tgg ctt atg cct gta atc cca gca ctt cag gag       702
Xaa Gly Arg Ala Arg Trp Leu Met Pro Val Ile Pro Ala Leu Gln Glu
        -15                    -10                    -5
gcc gan gca ggc gga tca cga ggt cag gag ttt gaa act agc ctg gcc       750
```

```
Ala Xaa Ala Gly Gly Ser Arg Gly Gln Glu Phe Glu Thr Ser Leu Ala
     1               5                   10
aac atg gag act gag gca gga gaa ttg ctt aaa ccc agg agg cgg agg    798
Asn Met Glu Thr Glu Ala Gly Glu Leu Leu Lys Pro Arg Arg Arg Arg
    15                  20                  25
ttg car tgaactgaga tcgcaccact gcactccagc ttgggcaaca gagcaagact    854
Leu Gln
30
ttgtctcgca aaaaaaaaaa a                                            875
```

```
<210> 362
<211> 531
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 70..366

<221> sig_peptide
<222> 70..108
<223> Von Heijne matrix
      score 3.5
      seq MHLLSNWANPASS/RR

<221> polyA_signal
<222> 496..501

<221> polyA_site
<222> 521..531

<400> 362
aagtggccat ggcggataca gcgactacag catcggcggc ggcggctagt gccgctagcg    60
cctcgagcg atg cac ctc ctt tcc aac tgg gca aac ccc gct tcc agc aga   111
           Met His Leu Leu Ser Asn Trp Ala Asn Pro Ala Ser Ser Arg
               -10                  -5                           1
cgt cct tct atg gcc gct tca ggc act tct tgg ata tca tcg acc ctc    159
Arg Pro Ser Met Ala Ala Ser Gly Thr Ser Trp Ile Ser Ser Thr Leu
            5                   10                  15
gca cac tct ttg tca ctg aga gac gtc tca gag agg ctg tgc agc tgc    207
Ala His Ser Leu Ser Leu Arg Asp Val Ser Glu Arg Leu Cys Ser Cys
        20                  25                  30
tgg agg act ata agc atg gga ccc tgc gcc cgg ggg tca cca atg aac    255
Trp Arg Thr Ile Ser Met Gly Pro Cys Ala Arg Gly Ser Pro Met Asn
    35                  40                  45
agc tct gga gtg cac aga aaa tca agc agg cta ttc tac atc cgg aca    303
Ser Ser Gly Val His Arg Lys Ser Ser Arg Leu Phe Tyr Ile Arg Thr
50                  55                  60                  65
cca atg aga aga tct tca tgc cat tta gaa tgt crg gtt ata ttc ctt    351
Pro Met Arg Arg Ser Ser Cys His Leu Glu Cys Xaa Val Ile Phe Leu
                70                  75                  80
ttg gga cgc caa ttg taaktgttac cttcaaagga tttccttttc taaaaaatta   406
Leu Gly Arg Gln Leu
             85
ttttaratgt ctaactttat gttattgctc acgggtattt gactgaattg ttgatttagg   466
ataagtcaat tcctggaggg aaattaccaa ataaaatgat atgtatttct taccacaaaa   526
aaaaa                                                               531
```

```
<210> 363
<211> 1244
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 70..366

<221> sig_peptide
<222> 70..108
<223> Von Heijne matrix
      score 3.5
      seq MHLLSNWANPASS/RR

<221> polyA_site
<222> 1233..1244

<400> 363
aagtggccat ggcggataca gcgactacag catcggcggc ggcggctagt gccgctagcg     60
cctcgagcg atg cac ctc ctt tcc aac tgg gca aac ccc gct tcc agc aga    111
          Met His Leu Leu Ser Asn Trp Ala Asn Pro Ala Ser Ser Arg
               -10              -5                    1
cgt cct tct atg gcc gct tca ggc act tct tgg ata tca tcg acc ctc       159
Arg Pro Ser Met Ala Ala Ser Gly Thr Ser Trp Ile Ser Ser Thr Leu
             5             10                  15
gca cac tct ttg tca ctg aga gac gtc tca gag agg ctg tgc agc tgc       207
Ala His Ser Leu Ser Leu Arg Asp Val Ser Glu Arg Leu Cys Ser Cys
         20              25                  30
tgg agg act ata agc atg gga ccc tgc gcc cgg ggg tca cca atg aac       255
Trp Arg Thr Ile Ser Met Gly Pro Cys Ala Arg Gly Ser Pro Met Asn
     35              40                  45
agc tct gga gtg cac aga aaa tca agc agg cta ttc tac atc cgg aca       303
Ser Ser Gly Val His Arg Lys Ser Ser Arg Leu Phe Tyr Ile Arg Thr
50              55                  60                  65
cca atg aga aga tct tca tgc cat tta raa tgt cag gtt ata ttc ctt       351
Pro Met Arg Arg Ser Ser Cys His Leu Xaa Cys Gln Val Ile Phe Leu
             70                  75                  80
ttg gga cgc caa ttg tagtcggtct tctcttgccc aaccagacac tggcatccac       406
Leu Gly Arg Gln Leu
             85
tgtcttctgg cagtggctga accagagcca caatgcctgt gtcaactatg caaaccgcaa     466
tgcraccaag ccttcacctg catccaagtt catccaggga tacctgggag ctgtcatcag     526
cgccgtctcc attgctgtgg gccttatktc ctggttcaga aagccaacaa gttcacccca     586
gccacccgcc ttctcatcca gaggtttgtg ccgttccctg ctgtagccag tgccaatatc     646
tgcaatgtgg tcctgatgcg gtacggggag ctggaggaag ggattgatgt cctggacagc     706
gatggcaacc tcgtgggctc ctccaagatc gcagcccgac acgccctgct ggagacggcg     766
ctgacgcgag tggtcctgcc catgcccatc ctggtgctac ccccgatcgt catgtccatg     826
ctggagaaga cggctctcct gcaggcacgc ccccggctgc tcctccctgt gcaaagcctc     886
gtgtgcctgg cagccttcgg cctggccctg ccgctggcca tcagcctctt cccgcaaatg     946
tcagagattg aaacatccca attagagccg gagatagccc aggccacgag cagccggaca    1006
gtggtgtaca acaaggggtt gtgagtgtgg tcagcggcct ggggacggag cactgtgcag    1066
ccggggagct gaggggcarg gccgtagact cacggctgca cctgcaggga gcagcacgcc    1126
aaccccagca gtcctgggcc ccctgggaga gtgctcaacc tacagtggag ggagactgac    1186
ccattcacat tttaacatag gcaagaggag ttctaacaca tttcgtacaa aaaaaaaa      1244


<210> 364
<211> 631
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 111..434

<221> sig_peptide
<222> 111..185
<223> Von Heijne matrix
      score 3.90000009536743
      seq WIAAVTIAAGTAA/IG

<221> polyA_site
<222> 618..631

<400> 364
aatcgcggag tcggtgcttt agtacgccgc tggcaccttt actctcgccg gccgcgcgaa     60
cccgtttgag ctcggtatcc tagtgcacac gccttgcaag cgacggcgcc atg agt     116
                                                         Met Ser
                                                         -25
ctg act tcc agt tcc agc gta cga gtt gaa tgg atc gca gca gtt acc     164
Leu Thr Ser Ser Ser Ser Val Arg Val Glu Trp Ile Ala Ala Val Thr
         -20             -15             -10
att gct gct ggg aca gct gca att ggt tat cta gct tac aaa aga ttt     212
Ile Ala Ala Gly Thr Ala Ala Ile Gly Tyr Leu Ala Tyr Lys Arg Phe
         -5              1               5
tat gtt aaa gat cat cga aat aaa gct atg ata aac ctt cac atc cag     260
Tyr Val Lys Asp His Arg Asn Lys Ala Met Ile Asn Leu His Ile Gln
10              15              20              25
aaa gac aac ccc aag ata gta cat gct ttt gac atg gag gat ttg gga     308
Lys Asp Asn Pro Lys Ile Val His Ala Phe Asp Met Glu Asp Leu Gly
         30              35              40
gat aaa gct gtg tac tgc cgt tgt tgg agg tcc aaa aag ttc cca ttc     356
Asp Lys Ala Val Tyr Cys Arg Cys Trp Arg Ser Lys Lys Phe Pro Phe
         45              50              55
tgt gat ggg gct cac aca aaa cat aac gaa gag act gga gac aat gtg     404
Cys Asp Gly Ala His Thr Lys His Asn Glu Glu Thr Gly Asp Asn Val
         60              65              70
ggc cct ctg atc atc aag aaa aaa gaa act taaatggaca cttttgatgc     454
Gly Pro Leu Ile Ile Lys Lys Lys Glu Thr
         75              80
tgcaaatcag cttgtcgtga agttacctga ttgtttaatt araatgacta ccacctctgt     514
ctgattcacc ttcgctggat tctaaatgtg gtatattgcm aactgcagct ttcacattta     574
tggcatttgt cttgttgaaa catcgtggtg cacatttgtt taaacaaaaa aaaaaaa     631


<210> 365
<211> 781
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 19..567

<221> sig_peptide
<222> 19..63
<223> Von Heijne matrix
      score 8.39999961853027
      seq AMWLLCVALAVLA/WG

<221> polyA_signal
```

```
<222> 749..754

<221> polyA_site
<222> 771..781

<400> 365
aagtgctgct tacccatc atg gaa gca atg tgg ctc ctg tgt gtg gcg ttg        51
                    Met Glu Ala Met Trp Leu Leu Cys Val Ala Leu
                    -15             -10              -5
gcg gtc ttg gca tgg ggc ttc ctc tgg gtt tgg gac tcc tca gaa cga        99
Ala Val Leu Ala Trp Gly Phe Leu Trp Val Trp Asp Ser Ser Glu Arg
            1               5               10
atg aag agt cgg gag cag gga aga cgg ctg gga gcc gaa agc cgg acc        147
Met Lys Ser Arg Glu Gln Gly Arg Arg Leu Gly Ala Glu Ser Arg Thr
        15              20              25
ctg ctg gtc ata gcg cac cct gac gat gaa gcc atg ttt ttt gct ccc        195
Leu Leu Val Ile Ala His Pro Asp Asp Glu Ala Met Phe Phe Ala Pro
        30              35              40
aca gtg cta ggc ttg gcc cgc cta agg cac tgg gtg tac ctg ctt tgc        243
Thr Val Leu Gly Leu Ala Arg Leu Arg His Trp Val Tyr Leu Leu Cys
45              50              55              60
ttc tct gca gga aat tac tac aat caa gga gag act cgt aag aaa gaa        291
Phe Ser Ala Gly Asn Tyr Tyr Asn Gln Gly Glu Thr Arg Lys Lys Glu
                65              70              75
ctt ttg car agc tgt gat gtt ttg ggg att cca ctc tcc agt gta atg        339
Leu Leu Gln Ser Cys Asp Val Leu Gly Ile Pro Leu Ser Ser Val Met
            80              85              90
att att gac aac agg gat ttc cca rat gac cca ggc atg cag tgg gac        387
Ile Ile Asp Asn Arg Asp Phe Pro Xaa Asp Pro Gly Met Gln Trp Asp
            95              100             105
aca rag cac gtg gcc ara gtc ctc ctt cag cac ata gaa gtg aat ggc        435
Thr Xaa His Val Ala Xaa Val Leu Leu Gln His Ile Glu Val Asn Gly
            110             115             120
atc aat ctg gtg gtg act ttc gat gca ggg gga rta agt ggc cac agc        483
Ile Asn Leu Val Val Thr Phe Asp Ala Gly Gly Xaa Ser Gly His Ser
125             130             135             140
aat cac att gct ctg tat gca gct gtg agg aag ctt gag ggc caa att        531
Asn His Ile Ala Leu Tyr Ala Ala Val Arg Lys Leu Glu Gly Gln Ile
            145             150             155
tgc aag ccc tgt ggc act gga caa gac ttt aag gaa tgagtgctgt            577
Cys Lys Pro Cys Gly Thr Gly Gln Asp Phe Lys Glu
            160             165
caatcagtgt gcctccacct tcaccatctt cttccccttc ctctcacttc cgtcatgtgt      637
tttatacaac tctcaaatct ttcttggaga aggaggatat acatacataa tatgaaatgt      697
gtttgttctt cacagtcacc cgatttttact gatatttatt tgcattttac caataaaaag    757
aaaatgcaag ctcaaaaaaa aaaa                                            781



<210> 366
<211> 931
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 19..312

<221> sig_peptide
<222> 19..63
<223> Von Heijne matrix
      score 8.39999961853027
```

seq AMWLLCVALAVLA/WG

<221> polyA_signal
<222> 896..901

<221> polyA_site
<222> 921..931

<400> 366

```
aagtgctgct tacccatc atg gaa gca atg tgg ctc ctg tgt gtg gcg ttg        51
                    Met Glu Ala Met Trp Leu Leu Cys Val Ala Leu
                    -15              -10                  -5
gcg gtc ttg gca tgg ggc ttc ctc tgg gtt tgg gac tcc tca gaa cga        99
Ala Val Leu Ala Trp Gly Phe Leu Trp Val Trp Asp Ser Ser Glu Arg
 1            .          5                 10
atg aag agt cgg gag cag gga rga cgg ctg gga gcc gaa agc cgg acc       147
Met Lys Ser Arg Glu Gln Gly Xaa Arg Leu Gly Ala Glu Ser Arg Thr
     15              20                 25
ctg ctg gtc ata gcg cac cct gac gat gaa gcc atg ttt ttt gct ccc       195
Leu Leu Val Ile Ala His Pro Asp Asp Glu Ala Met Phe Phe Ala Pro
     30              35                 40
aca gtg cta ggc ttg gcc cgc cta agg cac tgg gtg tac ctg ctt tgc       243
Thr Val Leu Gly Leu Ala Arg Leu Arg His Trp Val Tyr Leu Leu Cys
45              50                 55                 60
ttc tct gca gtt ttc cgt agg gag cta agt gaa tac acc gaa rgt ctt       291
Phe Ser Ala Val Phe Arg Arg Glu Leu Ser Glu Tyr Thr Glu Xaa Leu
                65                 70                 75
acc tct gaa ccc ctc ama gcc tagggacagg arcggccggc ttacctggtg          342
Thr Ser Glu Pro Leu Xaa Ala
                80
ggttggggga cgtcggcagc tcrcgtacta cgccagcagg attganganc acagaaacag     402
ttgchsttgg ttgtattcag tacctkcatt tccgttggga actccaccwg tacttgttat     462
kctgtggaac ttttttttat ttgtagaagg agcaagaata ttgaccttac tatatagcac     522
acgaaacaat ctatgctgta tcgtgcctgc tcaatcctta aagttaactt ctaatgatag     582
taaaaracct tcctgctgcc tttaaaatgc agcttgtgct aktaacatgc atgtgtcaaa     642
ttgaaraatt agacatagat gactaratar aaagtaattt tgtaggtaat tttaragttc     702
aactccaccc agctttcakt gaaggaacct ttcaaataat aratttttgc ttaccatara     762
raaaaratca aatgacaaag caaatattga ccattaagct ggaatatggt gataattgaa     822
cagttgtata aatgaaktaa ttgaattgta cacatacaat gggtgaattt tatggcatgt     882
caaagtatac ctcaataaag ctattttttt aaattgcmaa aaaaaaaaa                 931
```

<210> 367
<211> 849
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 64..612

<221> sig_peptide
<222> 64..234
<223> Von Heijne matrix
      score 3.79999995231628
      seq QLWLVMEFCGAGS/VT

<221> polyA_site
<222> 839..849

<400> 367

```
acatacgggc aagtttataa gggtcgtcat gtcaaaacgg gccagcttgc agccatcaag        60
gtt atg gat gtc aca ggg gat gaa gag gaa gaa atc aaa caa gaa att        108
    Met Asp Val Thr Gly Asp Glu Glu Glu Glu Ile Lys Gln Glu Ile
        -55                 -50                 -45
aac atg ttg aag aaa tat tct cat cac cgg aat att gct aca tac tat        156
Asn Met Leu Lys Lys Tyr Ser His His Arg Asn Ile Ala Thr Tyr Tyr
    -40                 -35                 -30
ggt gct ttt atc aaa aag aac cca cca ggc atg gat gac caa ctt tgg        204
Gly Ala Phe Ile Lys Lys Asn Pro Pro Gly Met Asp Asp Gln Leu Trp
    -25                 -20                 -15
ttg gtg atg gag ttt tgt ggt gct ggc tct gtc acc gac ctg atc aag        252
Leu Val Met Glu Phe Cys Gly Ala Gly Ser Val Thr Asp Leu Ile Lys
-10                 -5              1                   5
aac aca aaa ggt aac acg ttg aaa gag gag tgg att gca tac atc tgc        300
Asn Thr Lys Gly Asn Thr Leu Lys Glu Glu Trp Ile Ala Tyr Ile Cys
            10                  15                  20
msg gaa atc tta cgg ggg ctg art cac ctg cac cag cat aaa gtg att        348
Xaa Glu Ile Leu Arg Gly Leu Xaa His Leu His Gln His Lys Val Ile
            25                  30                  35
cat cga rat att aaa ggg caa aat gtc ttg ctg act gaa aat gca gaa        396
His Arg Xaa Ile Lys Gly Gln Asn Val Leu Leu Thr Glu Asn Ala Glu
        40                  45                  50
gtt aaa cta gtg gac ttt gga rtc akt gct cag ctt gat cga aca gtg        444
Val Lys Leu Val Asp Phe Gly Xaa Xaa Ala Gln Leu Asp Arg Thr Val
55                  60                  65                  70
ggc agg arg aat act ttc att gga act ccc tac tgg atg gca cca raa        492
Gly Arg Xaa Asn Thr Phe Ile Gly Thr Pro Tyr Trp Met Ala Pro Xaa
            75                  80                  85
gtt att gcc tgt gat gaa aac cca sat gcc aca tat gat ttc aar art        540
Val Ile Ala Cys Asp Glu Asn Pro Xaa Ala Thr Tyr Asp Phe Lys Xaa
            90                  95                  100
gac ttg tgg tct ttg ggt atc acc gcc att gaa atg gca gaa ggg ctc        588
Asp Leu Trp Ser Leu Gly Ile Thr Ala Ile Glu Met Ala Glu Gly Leu
            105                 110                 115
ccc ctc tct gtg aca tgc acc cca tgagagctct cttcctcatc ccccggaatc        642
Pro Leu Ser Val Thr Cys Thr Pro
            120                 125
cagcgcctcg gctgaagtct aagaagtggt caaaaaaatt ccagtcattt attgagagct       702
gcttggtaaa aaatcacagc cagcgaccag caacagaaca attgatgaag catccattta       762
tacgagacca acctaatgag cgacaggtcc gcattcaact caaggaccat attgatagaa       822
caaagaagaa gcgaggaaaa aaaaaaa                                           849
```

```
<210> 368
<211> 644
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 39..458

<221> sig_peptide
<222> 39..80
<223> Von Heijne matrix
      score 4.40000009536743
      seq FLTALLWRGRIPG/RQ

<221> polyA_signal
<222> 613..618
```

<221> polyA_site
<222> 633..644

<400> 368
```
agcggagacg cagagtcttg agcagcgcgn caggcacc atg ttc ctg act gcg ctc        56
                                            Met Phe Leu Thr Ala Leu
                                                                 -10
ctc tgg cgc ggc cgc att ccc ggc cgt cag tgg atc ggg aag cac cgg         104
Leu Trp Arg Gly Arg Ile Pro Gly Arg Gln Trp Ile Gly Lys His Arg
         -5                      1                   5
cgg ccg cgg ttc gtg tcg ttg cgc gcc aag cag aac atg atc cgc cgc         152
Arg Pro Arg Phe Val Ser Leu Arg Ala Lys Gln Asn Met Ile Arg Arg
     10              15              20
ctg gag atc gag gcg gag aac cat tac tgg ctg agc atg ccc tac atg         200
Leu Glu Ile Glu Ala Glu Asn His Tyr Trp Leu Ser Met Pro Tyr Met
25              30              35              40
acc cgg gag cag gag cgc ggc cac gcc gcg ttg cgc agg agg gag gcc         248
Thr Arg Glu Gln Glu Arg Gly His Ala Ala Leu Arg Arg Arg Glu Ala
                 45              50              55
ttc gag gcc ata aag gcg gcc gcc act tcc aag ttc ccc ccg cat aga         296
Phe Glu Ala Ile Lys Ala Ala Ala Thr Ser Lys Phe Pro Pro His Arg
             60              65              70
ttc att gcg gac cag ctc gac cat ctc aat vgt cac caa gaa atg gtc         344
Phe Ile Ala Asp Gln Leu Asp His Leu Asn Xaa His Gln Glu Met Val
         75              80              85
cta atc ctg agt cgt cac cct tgg att tta tgg atc acg gag ctg acc         392
Leu Ile Leu Ser Arg His Pro Trp Ile Leu Trp Ile Thr Glu Leu Thr
     90              95              100
atc ttt acc tgg tct gga ctg aaa aac tgt agc ttg tgt gaa aat gag         440
Ile Phe Thr Trp Ser Gly Leu Lys Asn Cys Ser Leu Cys Glu Asn Glu
105              110              115              120
ctt tgg acc agt ctt tat taaaacaaac aaacatgagt agtctgcata              488
Leu Trp Thr Ser Leu Tyr
                 125
tcgaatatct agagctctaa accccccaat acttaaaagt ctaattgctg tcctgtggtt      548
tcattagtct gataggaaga tagggattтc ctcagtcaca gatgatattt tgaaggaaag      608
ctgcaataaa gccacaatga tttgaaaaaa aaaaaa                               644
```

<210> 369
<211> 918
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 9..185

<221> sig_peptide
<222> 9..50
<223> Von Heijne matrix
      score 3.70000004768372
      seq AALVTVLFTGVRR/LH

<221> polyA_site
<222> 906..918

<400> 369
```
agctcagc atg gct gct tta gtg act gtt ctc ttc aca ggt gtc cgg agg        50
         Met Ala Ala Leu Val Thr Val Leu Phe Thr Gly Val Arg Arg
         -10                      -5
```

```
ctg cac tgc agc gcr scg ctt ggg cgg gcg gcc agt ggc grc tac agc        98
Leu His Cys Ser Ala Xaa Leu Gly Arg Ala Ala Ser Gly Xaa Tyr Ser
1               5                   10                  15
agg aac tgg ctg cca acc cct ccg gct acg ggc ccc tta ccg agc tcc       146
Arg Asn Trp Leu Pro Thr Pro Pro Ala Thr Gly Pro Leu Pro Ser Ser
            20                  25                  30
cag act ggt cat atg cgg atg gcc gcc ctg ctc ccc caa tgaaaggcca        195
Gln Thr Gly His Met Arg Met Ala Ala Leu Leu Pro Gln
            35                  40                  45
gcttcgaaaa aaagctgaaa gggagacktt tgcaaracra kttgtactgc tgtcacagga     255
aatggacgct ggattacaas catggcasct caggcagcar aakttgcagg aaraacaaag     315
gaagcaggaa aatgctctta aacccaaagg ggcttcactg aaaascccac ttccaaktca     375
ataaaaagca actcctgcct cccttcctca ccctgtctct ggatttcttt tctatcacct     435
aratgcttca tccagccara aaatagcctt cackktcccc atctgtcttc aragcaaaar     495
agctgggacm ccaaraacaa gctgttarat cactgcctgg gaggcttggc ttartactct     555
catctctggt tccattccag ttcagctaag tcttgcttta aaatttttac ctcctagctg     615
ggtgcggtgg ctcacgcctg taatcccagc actttgggag gctgaggcgg gcagatcaca     675
agatcaggag ttcgagacca gcctggccaa cccagcctgg tcaacatggt gaaaccctgt     735
ccctactaaa gatacaaaca attagccggg cgtggtgggg tgcgcttgta atcccagcta     795
ctcaggaggc tgaggcagga gaatcgctta aactcgggag gtagaggttg cagtgagcca     855
aggtcacacc attgcactcc aacctgggcg acaggcgag actctgtctc aaaaaaaaaa      915
aaa                                                                    918
```

```
<210> 370
<211> 472
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 14..316

<221> sig_peptide
<222> 14..121
<223> Von Heijne matrix
      score 5.19999980926514
      seq PLRLLNLLILIEG/SV

<221> polyA_signal
<222> 442..447

<221> polyA_site
<222> 458..471

<400> 370
attatataga gcc atg ggg cct tac aac gtg gca gtg cct tca gat gta        49
            Met Gly Pro Tyr Asn Val Ala Val Pro Ser Asp Val
            -35                 -30                 -25
tct cat gcc cgc ttt tat ttc tta ttt cat cga cca tta agg ctg tta       97
Ser His Ala Arg Phe Tyr Phe Leu Phe His Arg Pro Leu Arg Leu Leu
            -20                 -15                 -10
aat ctg ctc atc ctt att gag ggc agt gtc gtc ttc tat cag ctc tat      145
Asn Leu Leu Ile Leu Ile Glu Gly Ser Val Val Phe Tyr Gln Leu Tyr
            -5                  1                   5
tcc ttg ctg cgg tcg gag aag tgg aac cac aca ctt tcc atg gct ctc      193
Ser Leu Leu Arg Ser Glu Lys Trp Asn His Thr Leu Ser Met Ala Leu
            10                  15                  20
atc ctc ttc tgc aac tac tat gtt tta ttt aaa ctt ctc cgg gac aga      241
Ile Leu Phe Cys Asn Tyr Tyr Val Leu Phe Lys Leu Leu Arg Asp Arg
25                  30                  35                  40
```

```
wta kta tta ggc agg gca tac tcc tac cca ctc aac agt tat gaa ctc          289
Xaa Xaa Leu Gly Arg Ala Tyr Ser Tyr Pro Leu Asn Ser Tyr Glu Leu
            45                  50                  55
aag gca aac twa gct gcc tct caw caa tgagggagaa ctcagataaa              336
Lys Ala Asn Xaa Ala Ala Ser Xaa Gln
            60                  65
aatattttca tacgttctat ttttttcttg tgattttat aaatatttaa gatattttat      396
attttgtata ctattatgtt ttgaaagtcg ggaagagtaa gggatattaa atgtatccgt     456
aaacaaaaaa aaaaam                                                     472
```

<210> 371
<211> 1504
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 70..1092

<221> sig_peptide
<222> 70..234
<223> Von Heijne matrix
      score 4.09999990463257
      seq AVCAALLASHPTA/EV

<221> polyA_signal
<222> 1475..1480

<221> polyA_site
<222> 1493..1504

<400> 371
```
agaaatcgta ggacttccga aagcagcggc ggcgtttgct tcactgcttg gaagtgtgag          60
tgcgcgaag atg cga aag gtg gtt ttr att acc ggg gct agc agt ggc att         111
           Met Arg Lys Val Val Leu Ile Thr Gly Ala Ser Ser Gly Ile
               -55                  -50                  -45
ggc ctg gcc ctc tgc aag cgg ctg ctg gcg gaa gat gat gag ctt cat         159
Gly Leu Ala Leu Cys Lys Arg Leu Leu Ala Glu Asp Asp Glu Leu His
    -40                  -35                  -30
ctg tgt ttg gcg tgc agg aat atg agc aag gca gaa gct gtc tgt gct         207
Leu Cys Leu Ala Cys Arg Asn Met Ser Lys Ala Glu Ala Val Cys Ala
-25                  -20                  -15                  -10
gct ctg ctg gcc tct cac ccc act gct gag gtc acc att gtc cag gtg         255
Ala Leu Leu Ala Ser His Pro Thr Ala Glu Val Thr Ile Val Gln Val
              -5                   1                   5
gat gtc agc aac ctg cag tca ttc ttc cgg gcc tcc aag gaa ctt aag         303
Asp Val Ser Asn Leu Gln Ser Phe Phe Arg Ala Ser Lys Glu Leu Lys
             10                  15                  20
caa agg ttt cag aga tta gac tgt ata tat cta aat gct ggg atc atg         351
Gln Arg Phe Gln Arg Leu Asp Cys Ile Tyr Leu Asn Ala Gly Ile Met
    25                  30                  35
cct aat cca caa cta aat atc aaa gca ctt ttc ttt ggc ctc ttt tca         399
Pro Asn Pro Gln Leu Asn Ile Lys Ala Leu Phe Phe Gly Leu Phe Ser
40                  45                  50                  55
aga aaa gtg att cat atg ttc tcc aca gct gaa ggc ctg ctg acc cag         447
Arg Lys Val Ile His Met Phe Ser Thr Ala Glu Gly Leu Leu Thr Gln
            60                  65                  70
ggt gat aag atc act gct gat gga ctt cag gag gtg ttt gag acc aat         495
Gly Asp Lys Ile Thr Ala Asp Gly Leu Gln Glu Val Phe Glu Thr Asn
            75                  80                  85
```

```
gtc ttt ggc cat ttt atc ctg att cgg gaa ctg gag cct ctc ctc tgt        543
Val Phe Gly His Phe Ile Leu Ile Arg Glu Leu Glu Pro Leu Leu Cys
         90                  95                  100
cac agt gac aat cca tct cag ctc atc tgg aca tca tct cgc agt gca        591
His Ser Asp Asn Pro Ser Gln Leu Ile Trp Thr Ser Ser Arg Ser Ala
        105                 110                 115
agg aaa tct aat ttc agc ctc gag gac ttc cag cac agc aaa ggc aag        639
Arg Lys Ser Asn Phe Ser Leu Glu Asp Phe Gln His Ser Lys Gly Lys
120                 125                 130                 135
gaa ccc tac agc tct tcc aaa tat gcc act gac ctt ttg agt gtg gct        687
Glu Pro Tyr Ser Ser Ser Lys Tyr Ala Thr Asp Leu Leu Ser Val Ala
                140                 145                 150
ttg aac agg aac ttc aac cag cag ggt ctc tat tcc aat gtg gcc tgt        735
Leu Asn Arg Asn Phe Asn Gln Gln Gly Leu Tyr Ser Asn Val Ala Cys
            155                 160                 165
cca ggt aca gca ttg acc aat ttg aca tat gga att ctg cct ccg ttt        783
Pro Gly Thr Ala Leu Thr Asn Leu Thr Tyr Gly Ile Leu Pro Pro Phe
        170                 175                 180
ata tgg acg ctg ttg atg ccg gca ata ttg cta ctt cgc ttt ttt gca        831
Ile Trp Thr Leu Leu Met Pro Ala Ile Leu Leu Leu Arg Phe Phe Ala
        185                 190                 195
aat gca ttc act ttg aca cca tat aat gga aca gaa gct ctg gta tgg        879
Asn Ala Phe Thr Leu Thr Pro Tyr Asn Gly Thr Glu Ala Leu Val Trp
200                 205                 210                 215
ctt ttc cac caa aag cct gaa tct ctc aat cct ctg atc aaa tat ctg        927
Leu Phe His Gln Lys Pro Glu Ser Leu Asn Pro Leu Ile Lys Tyr Leu
                220                 225                 230
agt gcc acc act ggc ttt gga aga aat tac att atg acc cag aag atg        975
Ser Ala Thr Thr Gly Phe Gly Arg Asn Tyr Ile Met Thr Gln Lys Met
                235                 240                 245
gac cta gat gaa gac act gct gaa aaa ttt tat caa aag tta ctg gaa       1023
Asp Leu Asp Glu Asp Thr Ala Glu Lys Phe Tyr Gln Lys Leu Leu Glu
            250                 255                 260
ctg gaa aag cac att agg gtc act att caa aaa aca gat aat cag gcc       1071
Leu Glu Lys His Ile Arg Val Thr Ile Gln Lys Thr Asp Asn Gln Ala
        265                 270                 275
agg ctc agt ggc tca tgc cta taattccagc acttgggag gccaaggcag          1122
Arg Leu Ser Gly Ser Cys Leu
280                 285
aaggatcact tgagaccagg agttcaagac cagcctgaga aacatagtga gcccttgtct     1182
ctacaaaaag aaataaaaat aatagctggg tgtggtggca tgcgcatgta gtcccagcta     1242
ctcagaagga tgaggtggga ggatctcttg aggctgggag gcagaggttg cagtgagctg     1302
agattgtgcc actgcactcc agcctgggtg acagcgagac cctgtctcaa aatatgtata     1362
tatttaatat atatataaaa ccagagctga caatgacact ctggaacatt gcataccttc     1422
tgtacattct ggggtacatg gattctact gagttggata atatgcattt gtaataaact      1482
atgaactatg aaaaaaaaaa aa                                              1504
```

<210> 372
<211> 765
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 274..597

<221> sig_peptide
<222> 274..399
<223> Von Heijne matrix
      score 5.19999980926514

seq LLFDLVCHEFCQS/DD

```
<221> polyA_signal
<222> 731..736

<221> polyA_site
<222> 754..765

<400> 372
accaggaaca tccagctatt tatgatagca tttgcttcat tatgtcaagt tcaacaaatg        60
ttgacttgct ggtgaaggtg ggggaggttg tggacaagct ctttgatttg gatgagaaac       120
taatgttaag aatgggggct gctcagcctc tggaccaacc ccaggaagag                   180
tctgaagagc agccagtgtt tcggcttgtg ccctgtatac ttgaagctgc caaacaagta       240
cgttctgaaa atccagaatg gcttgatgtt tac atg cac att tta caa ctg ctt        294
                                     Met His Ile Leu Gln Leu Leu
                                                             -40
act aca gtg gat gat gga att caa gca att gta cat tgt cct gac act         342
Thr Thr Val Asp Asp Gly Ile Gln Ala Ile Val His Cys Pro Asp Thr
-35              -30                  -25                  -20
gga aaa gac att tgg aat tta ctt ttt gac ctg gtc tgc cat gaa ttc         390
Gly Lys Asp Ile Trp Asn Leu Leu Phe Asp Leu Val Cys His Glu Phe
                -15              -10                  -5
tgc cag tct gat gat cca gcc atc att ctt caa raa car aaa acr gtg         438
Cys Gln Ser Asp Asp Pro Ala Ile Ile Leu Gln Xaa Gln Lys Thr Val
            1               5                    10
cta gcc tct gtt ttt tca gtg ttg tct gcc atc tat gcc tca cag act         486
Leu Ala Ser Val Phe Ser Val Leu Ser Ala Ile Tyr Ala Ser Gln Thr
    15                  20                   25
gag caa gak tat cta aar ata raa aaa gga gac ggt ggc tca ggg agt         534
Glu Gln Xaa Tyr Leu Lys Ile Xaa Lys Gly Asp Gly Gly Ser Gly Ser
30                  35                   40                   45
aaa gga agg cca ktt gan caa aca gaa ktg ttc ctc tgc att tca aaa         582
Lys Gly Arg Pro Xaa Xaa Gln Thr Glu Xaa Phe Leu Cys Ile Ser Lys
                50                   55                   60
cct tct tcc ttt cta tagccctgtg gtggaagatt ttattaaaat cctacgtgaa         637
Pro Ser Ser Phe Leu
                65
gttgataagg cgcttgctga tgacttggaa aaaaacttcc caagtttgaa ggttcagact       697
taaaacctga attggaatta cttctgtaca agaaataaac tttattttc tcactgacaa        757
aaaaaaaa                                                                 765
```

```
<210> 373
<211> 1041
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 230..469

<221> sig_peptide
<222> 230..307
<223> Von Heijne matrix
      score 4.90000009536743
      seq VLCTNQVLITARA/VP

<221> polyA_signal
<222> 1004..1009

<221> polyA_site
```

```
<222> 1027..1040

<400> 373
aacttccaag ttgtagtgtt gttgttttca gcctgctgct gctgctgcta ttgcggctag        60
gggaaccgtc gtggggaagg atggtgtgcg aaaaatgtga aaagaaactt ggtactgtta       120
tcactccaga tacatggaaa gatggtgcta ggaataccac agaaagtggt ggaagaaagc       180
tgaatgaaaa taaagctttg acttcaaaaa aagccagaat tgatccata atg gaa gaa       238
                                                         Met Glu Glu
                                                             -25
ata agt tct cca ctt gta gaa ttt gta aaa gtt ttg tgc acc aac cag        286
Ile Ser Ser Pro Leu Val Glu Phe Val Lys Val Leu Cys Thr Asn Gln
            -20                 -15                 -10
gtt ctc att act gcc agg gct gtg cct aca aaa aag gca tct gtg cga        334
Val Leu Ile Thr Ala Arg Ala Val Pro Thr Lys Lys Ala Ser Val Arg
        -5                   1                   5
tgt gtg gaa aaa agg ttt tgg ata cca aaa act aca agc aaa cat ctg        382
Cys Val Glu Lys Arg Phe Trp Ile Pro Lys Thr Thr Ser Lys His Leu
10                  15                  20                  25
tct aga tgt att gat gga att tct ggc ttt cta aat gat ttt act ttc        430
Ser Arg Cys Ile Asp Gly Ile Ser Gly Phe Leu Asn Asp Phe Thr Phe
                30                  35                  40
tgc ctt gaa ttt tca agg cat aga tgt caa ctt aca gaa taacatgtkt        479
Cys Leu Glu Phe Ser Arg His Arg Cys Gln Leu Thr Glu
            45                  50
taagataatt aagtktaaac cagaraattt gattgttact cattttgctc tcatgtkcta       539
aaacagcaac agtgtaacta gtcttttgtt gtaaatggtt attttcctta taaaaatttt       599
aaaaactaag tggcaaattc catgaaaata tttctcagtt ctgtatgcac ttttatttaa       659
cattattcat ataattctcc ccccaccact ttatttataa atactgcaaa aktgaraagg       719
agataataaa tactttgctc tgaatttggc atccaaagtt aacatttctc ccctcactcc       779
cttgctggtg tcatagttat tagaatcagc agcctcttaa ctaattgcgg tttcatagga       839
tatataaatg tttcaagcca ttattgctga atggttcttt agttattaac ctagacccaa       899
atcaaagacc agttggattt atgatatttt ttatttgttc ttgcagccaa agtgccagtt       959
tctttaatat gtgaccaaga acacaaggag catccatatg gccaaataaa tacactgaat      1019
tttagaaaaa caaaaaaaaa ar                                               1041



<210> 374
<211> 1164
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 72..545

<221> sig_peptide
<222> 72..203
<223> Von Heijne matrix
      score 5.5
      seq ILFFTGWWIMIDA/AV

<221> polyA_site
<222> 1151..1162

<400> 374
aaagtcggcg tggacgtttg aggaagctgg gatacagcat ttaatgaaaa atttatgctt        60
aagaagtaaa a atg gca ggc ttc cta gat aat ttt cgt tgg cca gaa tgt       110
           Met Ala Gly Phe Leu Asp Asn Phe Arg Trp Pro Glu Cys
               -40                 -35
gaa tgt att gac tgg agt gag aga aga aat gct gtg gca tct gtt gtc       158
Glu Cys Ile Asp Trp Ser Glu Arg Arg Asn Ala Val Ala Ser Val Val
```

```
        -30                      -25                      -20
gca ggt ata ttg ttt ttt aca ggc tgg tgg ata atg att gat gca gct        206
Ala Gly Ile Leu Phe Phe Thr Gly Trp Trp Ile Met Ile Asp Ala Ala
-15                      -10                      -5                   1
gtg gtg tat cct aag cca gaa cag ttg aac cat gcc ttt cac aca tgt        254
Val Val Tyr Pro Lys Pro Glu Gln Leu Asn His Ala Phe His Thr Cys
              5                        10                       15
ggt gta ttt tcc aca ttg gct ttc ttc atg ata aat gct gta tcc aat        302
Gly Val Phe Ser Thr Leu Ala Phe Phe Met Ile Asn Ala Val Ser Asn
         20                       25                       30
gct cag gtg aga ggt gat agc tat gaa agc ggc tgt tta gga aga aca        350
Ala Gln Val Arg Gly Asp Ser Tyr Glu Ser Gly Cys Leu Gly Arg Thr
    35                       40                       45
ggt gct cga gtt tgg ctt ttc att ggt ttc atg ttg atg ttt ggg tca        398
Gly Ala Arg Val Trp Leu Phe Ile Gly Phe Met Leu Met Phe Gly Ser
50                       55                       60                   65
ctt att gct tcc atg tgg att ctt ttt ggt gca tat gtt acc caa aat        446
Leu Ile Ala Ser Met Trp Ile Leu Phe Gly Ala Tyr Val Thr Gln Asn
                   70                       75                       80
act gat gtt tat ccg gga cta gct gtg ttt ttt caa aat gca ctt ata        494
Thr Asp Val Tyr Pro Gly Leu Ala Val Phe Phe Gln Asn Ala Leu Ile
              85                       90                       95
ttt ttt agc act ctg atc tac aaa ttt gga aga acc gaa gag cta tgg        542
Phe Phe Ser Thr Leu Ile Tyr Lys Phe Gly Arg Thr Glu Glu Leu Trp
         100                      105                      110
acc tgagatcact tcttaagtca cattttcctt ttgttatatt ctgtttgtag            595
Thr
ataggttttt tatctctcag tacacattgc caaatggagt agattgtaca ttaaatgttt    655
tgtttctttta cattttatg ttctgagttt tgaaatagtt ttatgaaatt tctttatttt    715
tcattgcata gactgttaat atgtatataa tacaagacta tatgaattgg ataatgagta    775
tcagtttttt attcctgaga tttagaactt gatctactcc ctgagccagg gttacatcat    835
cttgtcattt tagaagtaac cactcttgtc tctctggctg ggcacggtgg ctcatgcctg    895
taatcccagc actttgggag gccgaggcgg gccgattgct tgaggtcaag tgtttgagac    955
cagcctggcc aacatggcga aaccccatct actaaaaata caaaaattag ccaggcatgg    1015
tggtgggtgc ctgtaatccc aactacctag gaggctgagg caggagaatc gcttgaaccc    1075
gggggggcaga ggttgyagtg agctgagttt gcgccactgc actctagcct ggggggagaaa  1135
gtgaaactcc ctctcaaaaa aaaaaaaamc                                     1164
```

```
<210> 375
<211> 1250
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 36..425

<221> sig_peptide
<222> 36..119
<223> Von Heijne matrix
      score 11.6000003814697
      seq LLLLVQLLRFLRA/DG

<221> polyA_signal
<222> 1215..1220

<221> polyA_site
<222> 1240..1250

<400> 375
```

```
atttcttccc cccgagctgg gcgtgcgcgg ccgca atg aac tgg gag ctg ctg        53
                                       Met Asn Trp Glu Leu Leu
                                                       -25
ctg tgg ctg ctg gtg ctg tgc gcg ctg ctc ctg ctc ttg gtg cag ctg      101
Leu Trp Leu Leu Val Leu Cys Ala Leu Leu Leu Leu Leu Val Gln Leu
    -20                 -15                 -10
ctg cgc ttc ctg agg gct gac ggc gac ctg acg cta cta tgg gcc gag      149
Leu Arg Phe Leu Arg Ala Asp Gly Asp Leu Thr Leu Leu Trp Ala Glu
    -5              1               5                   10
tgg cag gga cga cgc cca gaa tgg gag ctg act gat atg gtg gtg tgg      197
Trp Gln Gly Arg Arg Pro Glu Trp Glu Leu Thr Asp Met Val Val Trp
            15                  20                  25
gtg act gga gcc tcg agt gga att ggt gag gag ctg gct tac cag ttg      245
Val Thr Gly Ala Ser Ser Gly Ile Gly Glu Glu Leu Ala Tyr Gln Leu
            30                  35                  40
tct aaa cta gga gtt tct ctt gtg ctg tca gcc aga aga gtg cat gag      293
Ser Lys Leu Gly Val Ser Leu Val Leu Ser Ala Arg Arg Val His Glu
        45                  50                  55
ctg gaa agg gtg aaa aga aga tgc cta gag aat ggc aat tta aaa gaa      341
Leu Glu Arg Val Lys Arg Arg Cys Leu Glu Asn Gly Asn Leu Lys Glu
    60                  65                  70
aaa gat ata ctt gtt ttg ccc ctt gac ctg acc gac act ggt tcc cat      389
Lys Asp Ile Leu Val Leu Pro Leu Asp Leu Thr Asp Thr Gly Ser His
75                  80                  85                  90
gaa agc ggc tac caa agc tgt tct cca gga att tgg tagaatcgac          435
Glu Ser Gly Tyr Gln Ser Cys Ser Pro Gly Ile Trp
                95                  100
attctggtca acaatgtgga aatgtcccag cgttctctgt gcatggatac caacttggat    495
gtctacagaa agctaatgag agcttaacta cttagggacg gtgtccttga caaaatgtgk    555
kctgcctcac atgatcgaga ngaarcaagg aaagattgtt actgtgaata gcatcctggg    615
tatcatatct gtacctcttt ccattggata ctgtgctagc aagcatgctc tccggggktk    675
ktttaatggc cttcraacag aacttgccac atacccargt ataatagttt ctaacatttg    735
cccaggacct gtgcaatcaa atattgtgga aaattcccta gctggagaag tcacaaagac    795
tataggcaat aatggagacc agtcccacaa gatgacaacc agtcgttgtg tgcggctgat    855
gttaatcagc atggccaatg atttgaaaga agtttggatc tcagaacaac ctttcttgtt    915
agtaacatat ttgtggcaat acatgccaac ctgggcctgg tggataacca acaagatggg    975
gaagaaaagg attgagaact ttaagagtgg tgtggatgca gactcttctt attttaaaat   1035
ctttaagaca aaacatgact gaaaagagca cctgtacttt tcaagccact ggagggagaa   1095
atggaaaaca tgaaaacagc aatcttctta tgcttctgaa taatcaaaga ctaatttgtg   1155
attttacttt ttaatagata tgactttgct tccaacatgg aatgaaataa aaaataaaata   1215
ataaaagatt gccatgaatc ttgcaaaaaa aaaaa                              1250
```

```
<210> 376
<211> 947
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 155..751

<221> sig_peptide
<222> 155..340
<223> Von Heijne matrix
      score 3.70000004768372
      seq SILGIISVPLSIG/YC

<221> polyA_signal
<222> 912..917
```

```
<221> polyA_site
<222> 937..947

<400> 376
agtgaaaaga agatgcctag agaatggcaa tttaaaagaa aaagatatac ttgtttttgcc     60
ccttgacctg accgacactg gttcccatga agcggctacc aaagctgttc tccaggagtt    120
tggtagaatc gacattctgg tcaacaatgg tgga atg tcc cag cgt tct ctg tgc    175
                                     Met Ser Gln Arg Ser Leu Cys
                                                       -60
atg gat acc agc ttg gat gtc tac aga rag cta ata gag ctt aac tac    223
Met Asp Thr Ser Leu Asp Val Tyr Arg Xaa Leu Ile Glu Leu Asn Tyr
-55                 -50                 -45                 -40
tta ggg acg gtg tcc ttg aca aaa tgt gtt ctg cct cac atg atc gag    271
Leu Gly Thr Val Ser Leu Thr Lys Cys Val Leu Pro His Met Ile Glu
            -35                 -30                 -25
agg aag caa gga aag att gtt act gtg aat agc atc ctg ggt atc ata    319
Arg Lys Gln Gly Lys Ile Val Thr Val Asn Ser Ile Leu Gly Ile Ile
            -20                 -15                 -10
tct gta cct ctt tcc att gga tac tgt gct agc aag cat gct ctc cgg    367
Ser Val Pro Leu Ser Ile Gly Tyr Cys Ala Ser Lys His Ala Leu Arg
        -5                   1                   5
ggt ttt ttt aat ggc ctt cga aca gaa ctt gcc aca tac cca ggt ata    415
Gly Phe Phe Asn Gly Leu Arg Thr Glu Leu Ala Thr Tyr Pro Gly Ile
10                  15                  20                  25
ata gtt tct aac att tgc cca gga cct gtg caa tca aat att gtg gaa    463
Ile Val Ser Asn Ile Cys Pro Gly Pro Val Gln Ser Asn Ile Val Glu
                    30                  35                  40
aat tcc cta gct gga gaa gtc aca aaa act ata ggc aat aat gga aac    511
Asn Ser Leu Ala Gly Glu Val Thr Lys Thr Ile Gly Asn Asn Gly Asn
            45                  50                  55
cag tcc cac aag atg aca acc agt cgt tgt gtg cgg ctg atg tta atc    559
Gln Ser His Lys Met Thr Thr Ser Arg Cys Val Arg Leu Met Leu Ile
        60                  65                  70
agc atg gcc aat gat ttg aaa gaa gtt tgg atc tca gaa caa cct ttc    607
Ser Met Ala Asn Asp Leu Lys Glu Val Trp Ile Ser Glu Gln Pro Phe
        75                  80                  85
ttg tta gta aca tat ttg tgg caa tac atg cca acc tgg gcc tgg tgg    655
Leu Leu Val Thr Tyr Leu Trp Gln Tyr Met Pro Thr Trp Ala Trp Trp
90                  95                  100                 105
ata acc aac aag atg ggg aag aaa agg att gag aac ttt aag agt ggt    703
Ile Thr Asn Lys Met Gly Lys Lys Arg Ile Glu Asn Phe Lys Ser Gly
            110                 115      .           120
gtg gat gcm rac tct tct tat ttt aaa atc ttt aag aca aaa cat gac    751
Val Asp Ala Xaa Ser Ser Tyr Phe Lys Ile Phe Lys Thr Lys His Asp
            125                 130                 135
tgaaaaganc acctgtactt ttcaagccac tggagggaga aatggaaaac atgaaaacag    811
caatcttctt atgcttctga ataatcaaag actaatttgt gatttttactt tttaatagat    871
atgacttttgc ttccaacatg grrtgaaata aaaaataaat aataaaagat tgccatgrrt    931
cttgcaaaaa aaaaaa                                                    947
```

```
<210> 377
<211> 621
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 46..585

<221> sig_peptide
```

```
<222> 46..120
<223> Von Heijne matrix
      score 6.30000019073486
      seq AFSLSVMAALTFG/CF

<221> polyA_signal
<222> 584..589

<221> polyA_site
<222> 606..619

<400> 377
aactgggtgt gcgtrtggag tccggactcg tgggagacga tcgcg atg aac acg gtg     57
                                                     Met Asn Thr Val
                                                         -25
ctg tcg cgg gcg aac tca ctg ttc gcc ttc tcg ctg agc gtg atg gcs    105
Leu Ser Arg Ala Asn Ser Leu Phe Ala Phe Ser Leu Ser Val Met Ala
    -20              -15              -10
gcg ctc acc ttc ggc tgc ttc atc ayy acc gcc ttc aaa gac agg agc    153
Ala Leu Thr Phe Gly Cys Phe Ile Xaa Thr Ala Phe Lys Asp Arg Ser
-5               1               5               10
gtc ccg gtg cgg ctg cac gtc tcg cga atc atg cta aaa aat gta gaa    201
Val Pro Val Arg Leu His Val Ser Arg Ile Met Leu Lys Asn Val Glu
        15              20              25
gat ttc act gga cct aga gaa aga agt gat ctg gga ttt atc aca ttt    249
Asp Phe Thr Gly Pro Arg Glu Arg Ser Asp Leu Gly Phe Ile Thr Phe
        30              35              40
gat ata act gct gat cta gag aat ata ttt gat tgg aat gtt aag cag    297
Asp Ile Thr Ala Asp Leu Glu Asn Ile Phe Asp Trp Asn Val Lys Gln
    45              50              55
ttg ttt ctt tat tta tca gca gaa tat tca aca aaa aat aat gct ctg    345
Leu Phe Leu Tyr Leu Ser Ala Glu Tyr Ser Thr Lys Asn Asn Ala Leu
60              65              70              75
aac caa ktt gtc cta tgg gac aag att gtt ttg aga ggt gat aat ccg    393
Asn Gln Xaa Val Leu Trp Asp Lys Ile Val Leu Arg Gly Asp Asn Pro
            80              85              90
aag ctg ctg ctg aaa gat atg aaa aca aaa tat ttt ttc ttt gac gat    441
Lys Leu Leu Leu Lys Asp Met Lys Thr Lys Tyr Phe Phe Phe Asp Asp
            95              100             105
gga aat ggt ctc wag gga aac agg aat gtc act ttg acc ctg tct tgg    489
Gly Asn Gly Leu Xaa Gly Asn Arg Asn Val Thr Leu Thr Leu Ser Trp
        110             115             120
aac gtc gta cca aat gct gga att cta cct ctt gtg aca gga tca gga    537
Asn Val Val Pro Asn Ala Gly Ile Leu Pro Leu Val Thr Gly Ser Gly
        125             130             135
cac gta tct gtc cca ttt cca gat aca tat gaa ata acg aag agt tat    585
His Val Ser Val Pro Phe Pro Asp Thr Tyr Glu Ile Thr Lys Ser Tyr
140             145             150             155
taaattattc tgaatttgaa acaaaaaaaa aaaahm                             621
```

```
<210> 378
<211> 52
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -20..-1

<400> 378
```

```
Met Pro Ser Val Asn Ser Ala Gly Leu Cys Val Leu Gln Leu Thr Thr
-20                -15                -10                        -5
Ala Val Thr Ser Ala Phe Leu Leu Ala Lys Val Asn Pro Phe Glu Xaa
                1               5                        10
Phe Leu Ser Arg Gly Phe Trp Leu Cys Ala Ala His His Phe Ile His
        15               20                        25
Pro Cys Leu Asp
    30
```

<210> 379
<211> 193
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -23..-1

<400> 379
```
Met Val Val Leu Arg Ala Gly Lys Lys Thr Phe Leu Pro Pro Leu Xaa
            -20               -15                        -10
Arg Ala Phe Ala Cys Arg Gly Cys Gln Leu Ala Pro Glu Arg Gly Ala
        -5                    1               5
Glu Arg Arg Asp Thr Ala Pro Ser Gly Val Ser Arg Phe Cys Pro Pro
10                    15                    20                    25
Arg Lys Ser Cys His Asp Trp Ile Gly Pro Pro Asp Lys Tyr Ser Asn
                30                    35                    40
Leu Arg Pro Val His Phe Tyr Ile Pro Glu Asn Glu Ser Pro Leu Glu
            45                    50                    55
Gln Lys Leu Arg Lys Leu Arg Gln Glu Thr Gln Glu Trp Asn Gln Gln
        60                    65                    70
Phe Trp Ala Asn Gln Asn Leu Thr Phe Ser Lys Glu Lys Glu Glu Phe
    75                    80                    85
Ile His Ser Arg Leu Lys Thr Lys Gly Leu Gly Leu Arg Thr Glu Ser
90                    95                    100                    105
Gly Gln Lys Ala Thr Leu Asn Ala Glu Glu Met Ala Asp Phe Tyr Lys
                110                    115                    120
Glu Phe Leu Ser Lys Asn Phe Gln Lys His Met Tyr Tyr Asn Arg Asp
                125                    130                    135
Trp Tyr Lys Arg Asn Phe Ala Ile Thr Phe Phe Met Gly Lys Val Ala
            140                    145                    150
Leu Glu Arg Ile Trp Asn Lys Leu Lys Gln Lys Gln Lys Lys Arg Ser
    155                    160                    165
Asn
170
```

<210> 380
<211> 82
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -14..-1

<400> 380
```
Met Ala Phe Thr Leu Xaa Ser Leu Leu Gln Ala Ala Leu Leu Cys Val
                -10               -5                    1
```

```
Asn Ala Ile Ala Val Leu His Glu Glu Arg Phe Leu Lys Asn Ile Gly
          5                   10                  15
Trp Gly Thr Asp Gln Gly Ile Gly Gly Phe Gly Glu Glu Pro Gly Ile
    20                  25                  30
Lys Ser Xaa Xaa Met Xaa Leu Ile Arg Ser Val Arg Thr Val Met Arg
35                  40                  45                  50
Val Pro Leu Ile Ile Val Asn Ser Ile Ala Ile Val Leu Leu Leu Leu
                55                  60                  65
Phe Gly
```

<210> 381
<211> 198
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 381
```
Met Pro Val Pro Ala Leu Cys Leu Leu Trp Ala Leu Ala Met Val Thr
    -20                 -15                 -10
Arg Pro Ala Ser Ala Ala Pro Met Gly Gly Pro Glu Leu Ala Gln His
    -5              1               5                   10
Glu Glu Leu Thr Leu Leu Phe His Gly Thr Leu Gln Leu Gly Gln Ala
              15                  20                  25
Leu Asn Gly Val Tyr Arg Thr Thr Glu Gly Arg Leu Thr Lys Ala Arg
          30                  35                  40
Asn Ser Leu Gly Leu Tyr Gly Arg Thr Ile Glu Leu Leu Gly Gln Glu
          45                  50                  55
Val Ser Arg Gly Arg Asp Ala Ala Gln Glu Leu Arg Ala Ser Leu Leu
60                  65                  70                  75
Glu Thr Gln Met Glu Glu Asp Ile Leu Xaa Leu Gln Ala Xaa Ala Thr
                80                  85                  90
Ala Glu Val Leu Gly Glu Val Ala Gln Ala Gln Lys Val Leu Arg Asp
              95                  100                 105
Ser Val Gln Arg Leu Xaa Xaa Gln Leu Xaa Xaa Ala Trp Leu Gly Pro
          110                 115                 120
Ala Tyr Arg Lys Phe Glu Val Leu Lys Ala Pro Pro Xaa Lys Gln Asn
          125                 130                 135
His Ile Leu Trp Ala Leu Thr Gly His Val Xaa Arg Gln Xaa Arg Glu
140                 145                 150                 155
Met Val Ala Gln Gln Xaa Xaa Leu Xaa Gln Ile Gln Glu Lys Leu His
                160                 165                 170
Thr Ala Ala Leu Pro Ala
          175
```

<210> 382
<211> 160
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -55..-1

<400> 382
```
Met Asp Lys Leu Lys Lys Val Leu Ser Gly Gln Asp Thr Glu Asp Arg
```

```
-55                      -50                      -45                      -40
Ser Gly Leu Ser Glu Val Val Glu Ala Ser Ser Leu Ser Trp Ser Thr
                -35                      -30                      -25
Arg Ile Lys Gly Phe Ile Ala Cys Phe Ala Ile Gly Ile Leu Cys Ser
                -20                      -15                      -10
Leu Leu Gly Thr Val Leu Leu Trp Val Pro Arg Lys Gly Leu His Leu
                -5                       1                        5
Phe Ala Val Phe Tyr Thr Phe Gly Asn Ile Ala Ser Ile Gly Ser Thr
10                       15                       20                       25
Ile Phe Leu Met Gly Pro Val Lys Gln Leu Lys Arg Met Phe Glu Pro
                30                       35                       40
Thr Arg Leu Ile Ala Thr Ile Met Val Leu Leu Cys Phe Ala Leu Thr
                45                       50                       55
Leu Cys Ser Ala Phe Trp Trp His Asn Lys Gly Leu Ala Leu Ile Phe
                60                       65                       70
Cys Ile Leu Gln Ser Leu Ala Leu Thr Trp Tyr Ser Leu Ser Phe Ile
     75                       80                       85
Pro Phe Ala Arg Asp Ala Val Lys Xaa Cys Phe Ala Val Cys Leu Ala
90                       95                       100                      105
```

```
<210> 383
<211> 108
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -18..-1

<400> 383
Met Lys Ala Leu Cys Leu Leu Leu Leu Pro Val Leu Gly Leu Leu Val
                -15                      -10                      -5
Ser Ser Lys Thr Leu Cys Ser Met Glu Glu Ala Ile Asn Glu Arg Ile
                1                        5                        10
Gln Glu Val Ala Gly Ser Leu Ile Phe Arg Ala Ile Ser Ser Ile Gly
15                       20                       25                       30
Arg Gly Ser Glu Ser Val Thr Ser Arg Gly Asp Leu Ala Thr Cys Pro
                35                       40                       45
Arg Gly Phe Ala Val Thr Gly Cys Thr Cys Gly Ser Ala Cys Gly Ser
                50                       55                       60
Trp Asp Val Arg Ala Glu Thr Thr Cys His Cys Gln Cys Ala Gly Met
                65                       70                       75
Asp Trp Thr Gly Ala Arg Cys Cys Arg Val Gln Pro
     80                       85                       90
```

```
<210> 384
<211> 64
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -22..-1

<400> 384
Met Ile Ser Arg Gln Leu Arg Ser Leu Ser Cys Leu Cys Pro Ala Leu
     -20                      -15                      -10
Phe Pro Gly Thr Ser Ser Phe Ile Val Ala Leu Ser Ser Pro Ala Asp
```

```
       -5                    1              5                        10
Leu Tyr Ile Pro Xaa Arg Xaa Arg Ser Asp Glu Leu Val Phe Glu Ser
                    15                  20                  25
Gln Lys Gly Ser Ala Met Glu Leu Ala Val Ile Thr Val Xaa Gly Val
                30                  35                  40
```

```
<210> 385
<211> 27
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -15..-1

<400> 385
Met Gly Phe Leu Xaa Leu Met Thr Leu Thr Thr His Val His Ser Ser
-15                 -10                 -5                      1
Ala Lys Pro Asn Glu Gln Pro Trp Leu Leu Asn
                5                   10
```

```
<210> 386
<211> 186
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 386
Met Ser Pro Ser Gly Arg Leu Cys Leu Leu Thr Ile Val Gly Leu Ile
        -20                 -15                 -10
Leu Pro Thr Arg Gly Gln Thr Leu Lys Asp Thr Thr Ser Ser Ser Ser
-5                   1                  5                   10
Ala Asp Ser Thr Ile Met Asp Ile Gln Val Pro Thr Arg Ala Pro Asp
            15                  20                  25
Ala Val Tyr Thr Glu Leu Gln Pro Thr Ser Pro Thr Pro Thr Trp Pro
        30                  35                  40
Ala Asp Glu Thr Pro Gln Pro Gln Thr Gln Thr Gln Gln Leu Glu Gly
        45                  50                  55
Thr Asp Gly Pro Leu Val Thr Asp Pro Glu Thr His Xaa Ser Xaa Lys
60                  65                  70                  75
Ala Ala His Pro Thr Asp Asp Thr Thr Thr Leu Ser Glu Arg Pro Ser
            80                  85                  90
Pro Ser Thr Xaa Val His Xaa Arg Pro Xaa Xaa Pro Ser Xaa His Leu
        95                  100                 105
Val Phe Met Arg Met Thr Pro Ser Ser Met Met Asn Thr Pro Ser Gly
        110                 115                 120
Asn Xaa Gly Cys Trp Ser Gln Leu Cys Cys Ser Ser Gln Ala Ser Ser
    125                 130                 135
Ser Ser Pro Val Ala Ser Ala Gly Ser Cys Pro Gly Tyr Ala Gly Ile
140                 145                 150                 155
Ile Ala Gly Glu Ser Ile Arg Asn Arg Ser
                160                 165
```

```
<210> 387
<211> 179
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -26..-1

<400> 387
Met Glu Thr Gly Ala Leu Arg Arg Pro Gln Leu Leu Pro Leu Leu Leu
    -25                   -20                   -15
Leu Leu Cys Gly Pro Ser Gln Asp Gln Cys Arg Pro Val Leu Gln Asn
-10                   -5                    1               5
Leu Leu Gln Ser Pro Gly Leu Thr Trp Ser Leu Glu Val Pro Thr Gly
              10                  15                  20
Arg Glu Gly Lys Glu Gly Gly Asp Arg Gly Pro Gly Leu Xaa Gly Ala
          25                  30                  35
Thr Pro Ala Arg Ser Pro Gln Gly Lys Glu Met Gly Arg Gln Arg Th_
    40                  45                  50
Arg Lys Val Lys Gly Pro Ala Trp Xaa His Thr Ala Asn Gln Glu Leu
55                  60                  65                  70
Asn Arg Met Arg Ser Leu Ser Ser Gly Ser Val Pro Val Gly His Leu
              75                  80                  85
Glu Gly Gly Thr Val Lys Leu Gln Lys Asp Thr Gly Leu His Ser Cys
              90                  95                  100
Xaa Asp Gly Met Ala Ser Leu Glu Gly Thr Pro Ala Ser Val Leu Ala
          105                 110                 115
Asp Ala Cys Pro Gly Phe His Asp Val Xaa Val Gln Xaa Ala Leu Phe
    120                 125                 130
Gly Leu Ser Gly Xaa Xaa Leu Trp Leu Lys Thr His Phe Cys Leu Ser
135                 140                 145                 150
Ile Xaa Leu




<210> 388
<211> 150
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -55..-1

<400> 388
Met Ala Thr Thr Val Pro Asp Gly Cys Arg Asn Gly Leu Lys Ser Lys
-55                   -50                   -45                 -40
Tyr Tyr Arg Leu Cys Asp Lys Ala Glu Ala Trp Gly Ile Val Leu Glu
              -35                 -30                 -25
Thr Val Ala Thr Ala Gly Val Val Thr Ser Val Ala Phe Met Leu Thr
              -20                 -15                 -10
Leu Pro Ile Leu Val Cys Lys Val Gln Asp Ser Asn Arg Arg Lys Met
          -5                  1                   5
Leu Pro Thr Gln Phe Leu Phe Leu Leu Gly Val Leu Gly Ile Phe Gly
10                  15                  20                  25
Leu Thr Phe Ala Phe Ile Ile Gly Leu Asp Gly Ser Thr Gly Pro Thr
              30                  35                  40
Arg Phe Phe Leu Phe Gly Ile Leu Phe Ser Ile Cys Phe Ser Cys Leu
              45                  50                  55
Leu Ala His Ala Val Ser Leu Thr Lys Leu Val Arg Gly Arg Lys Ala
              60                  65                  70
```

```
Pro Phe Pro Val Gly Asp Ser Gly Ser Gly Arg Gly Leu Gln Pro Ser
    75                  80                  85
Pro Gly Cys Tyr Arg Tyr
90                  95
```

<210> 389
<211> 236
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -31..-1

```
<400> 389
Met Leu Ser Lys Gly Leu Lys Arg Lys Arg Glu Glu Glu Glu Glu Lys
    -30                 -25                 -20
Glu Pro Leu Ala Val Asp Ser Trp Trp Leu Asp Pro Gly His Ala Ala
-15                 -10                 -5                  1
Val Ala Gln Ala Pro Pro Ala Val Ala Ser Ser Ser Leu Phe Asp Leu
            5                   10                  15
Ser Val Leu Lys Leu His His Ser Leu Gln Xaa Ser Xaa Pro Asp Leu
        20                  25                  30
Arg His Leu Val Leu Val Xaa Asn Thr Leu Arg Arg Ile Gln Ala Ser
        35                  40                  45
Met Ala Pro Ala Ala Ala Leu Pro Pro Val Pro Thr Pro Pro Ala Ala
50                  55                  60                  65
Pro Xaa Val Ala Asp Asn Leu Leu Ala Ser Ser Asp Ala Ala Leu Ser
        70                  75                  80
Ala Ser Met Ala Xaa Leu Leu Glu Asp Leu Ser His Ile Glu Gly Leu
        85                  90                  95
Ser Gln Ala Pro Gln Pro Leu Ala Asp Glu Gly Pro Pro Gly Arg Ser
        100                 105                 110
Ile Gly Gly Xaa Pro Pro Xaa Leu Gly Ala Leu Asp Leu Leu Gly Pro
    115                 120                 125
Ala Thr Gly Cys Leu Leu Asp Asn Gly Leu Glu Gly Leu Phe Glu Asp
130                 135                 140                 145
Ile Asp Thr Ser Met Tyr Asp Asn Glu Leu Trp Ala Pro Ala Ser Glu
            150                 155                 160
Gly Leu Lys Pro Gly Pro Glu Asp Gly Pro Gly Lys Glu Glu Ala Pro
            165                 170                 175
Glu Leu Asp Glu Ala Glu Leu Asp Tyr Leu Met Asp Val Leu Val Gly
        180                 185                 190
Thr Gln Ala Leu Glu Arg Pro Pro Gly Pro Gly Arg
    195                 200                 205
```

<210> 390
<211> 149
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -100..-1

```
<400> 390
Met Glu Thr Leu Tyr Arg Val Pro Phe Leu Val Leu Glu Cys Pro Asn
-100                -95                 -90                 -85
```

```
Leu Lys Leu Lys Lys Pro Pro Trp Leu His Met Pro Ser Ala Met Thr
              -80                   -75                   -70
Val Tyr Ala Leu Val Val Val Ser Tyr Phe Leu Ile Thr Gly Gly Ile
          -65                   -60                   -55
Ile Tyr Asp Val Ile Val Glu Pro Pro Ser Val Gly Ser Met Thr Asp
          -50                   -45                   -40
Glu His Gly His Gln Arg Pro Val Ala Phe Leu Ala Tyr Arg Val Asn
      -35                   -30                   -25
Gly Gln Tyr Ile Met Glu Gly Leu Ala Ser Ser Phe Leu Phe Thr Met
-20                   -15                   -10                   -5
Gly Gly Leu Gly Phe Ile Ile Leu Asp Gly Ser Asn Ala Pro Asn Ile
1                     5                     10
Pro Lys Leu Asn Arg Phe Leu Leu Phe Ile Gly Phe Val Cys Val
      15                  20                  25
Leu Xaa Ser Phe Xaa Xaa Ala Arg Val Phe Met Arg Met Lys Leu Pro
    30              35                  40
Gly Tyr Leu Met Gly
45
```

```
<210> 391
<211> 69
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -49..-1

<400> 391
Met Pro Phe His Phe Pro Phe Leu Gly Phe Val Cys Leu His Leu His
              -45                   -40                   -35
Leu Thr Pro Cys Leu Thr Val Pro Arg Arg Pro Leu Phe Leu Leu Leu
          -30               '     -25                   -20
His Leu Cys Pro His Leu Pro Phe Leu Leu Leu Leu Ser Cys Val Gly
          -15                   -10                   -5
Xaa Xaa Pro Ser Cys Leu Pro Ser Ser Ser Thr Cys Val Ser Leu His
    1           5                     10                    15
Phe Phe Ile Pro Asp
              20
```

```
<210> 392
<211> 241
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -30..-1

<400> 392
Met Gly Thr Ala Ser Arg Ser Asn Ile Ala Arg His Leu Gln Thr Asn
-30                   -25                   -20                   -15
Leu Ile Leu Phe Cys Val Gly Ala Val Gly Ala Cys Thr Leu Ser Val
              -10                   -5
Thr Gln Pro Trp Tyr Leu Glu Val Asp Tyr Thr His Glu Ala Val Thr
          5                     10                    15
Ile Lys Cys Thr Phe Ser Ala Thr Gly Cys Pro Ser Glu Gln Pro Thr
      20                  25                  30
```

```
Cys Leu Trp Phe Arg Tyr Gly Ala His Gln Pro Glu Asn Leu Cys Leu
35                  40                  45                  50
Asp Gly Cys Lys Ser Glu Ala Xaa Lys Phe Thr Val Arg Glu Ala Leu
                55                  60                  65
Lys Glu Asn Gln Val Ser Leu Thr Val Asn Arg Val Thr Ser Asn Asp
            70                  75                  80
Ser Ala Ile Tyr Ile Cys Gly Ile Ala Phe Pro Ser Val Pro Glu Ala
        85                  90                  95
Arg Ala Lys Gln Thr Gly Gly Gly Thr Thr Leu Val Val Arg Glu Ile
    100                 105                 110
Lys Leu Leu Ser Lys Glu Leu Arg Ser Phe Leu Thr Ala Leu Val Ser
115                 120                 125                 130
Leu Leu Ser Val Tyr Val Thr Gly Val Cys Val Ala Phe Ile Leu Leu
            135                 140                 145
Ser Lys Ser Lys Ser Asn Pro Leu Arg Asn Lys Glu Ile Lys Glu Asp
            150                 155                 160
Ser Gln Lys Lys Lys Ser Ala Arg Arg Ile Phe Gln Glu Ile Ala Gln
        165                 170                 175
Glu Leu Tyr His Lys Arg His Val Glu Thr Asn Gln Gln Ser Glu Lys
    180                 185                 190
Asp Asn Asn Thr Tyr Glu Asn Arg Arg Val Leu Ser Asn Tyr Glu Arg
195                 200                 205                 210
Pro
```

```
<210> 393
<211> 47
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -30..-1

<400> 393
Met Asn Cys Asn Val Val Ser Glu Arg Gly Lys Trp Leu Glu Val Glu
-30                 -25                 -20                 -15
Cys Ser Leu Met Thr Cys Thr Thr Leu Ile Asn Ala Ser Ala Ile Ser
                -10                 -5                  1
Thr Asn Thr Leu Thr Asp Met Gly Ser Phe Asp Arg Arg Glu Ser
            5                   10                  15
```

```
<210> 394
<211> 65
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -28..-1

<400> 394
Met Ala Phe Gly Leu Gln Met Phe Ile Gln Arg Lys Phe Pro Tyr Pro
                -25                 -20                 -15
Leu Gln Trp Ser Leu Leu Val Ala Val Val Ala Gly Ser Val Val Ser
            -10                 -5                  1
Tyr Gly Val Thr Arg Val Glu Ser Glu Lys Cys Asn Asn Leu Trp Leu
5                   10                  15                  20
Phe Leu Glu Thr Gly Gln Leu Pro Lys Asp Arg Ser Thr Asp Gln Xaa
```

```
                    25                    30                    35
Ser
```

```
<210> 395
<211> 73
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -24..-1

<400> 395
Met Thr Cys Trp Met Leu Pro Pro Ile Ser Phe Leu Ser Tyr Leu Pro
            -20                   -15                   -10
Leu Trp Leu Gly Pro Ile Trp Pro Cys Ser Gly Ser Thr Leu Gly Lys
            -5                    1                     5
Pro Asp Pro Gly Val Trp Pro Ser Leu Phe Arg Pro Trp Asp Ala Ala
    10                    15                    20
Ser Pro Gly Asn Tyr Ala Leu Ser Arg Gly Xaa Asn Xaa Tyr Xaa Xaa
25                    30                    35                    40
Trp Gly Gln Gly Thr His Ser Ser Leu
                45
```

```
<210> 396
<211> 60
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -18..-1

<400> 396
Met Pro Cys Pro Thr Trp Thr Cys Leu Lys Ser Phe Pro Ser Pro Thr
            -15                   -10                   -5
Ser Ser His Ala Ser Ser Leu His Leu Pro Pro Ser Cys Thr Arg Leu
    1                     5                     10
Thr Leu Thr Gln Thr Leu Arg Thr Gly Met His Leu Ser Arg Ala Leu
15                    20                    25                    30
Gln Gly Thr Leu Thr Arg Leu Gln Ser Thr Pro Ala
                35                    40
```

```
<210> 397
<211> 192
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -93..-1

<400> 397
Met Ala Glu Leu Gly Leu Asn Glu His His Gln Asn Glu Val Ile Asn
            -90                   -85                   -80
Tyr Met Arg Phe Ala Arg Ser Lys Arg Gly Leu Arg Leu Lys Thr Val
```

```
         -75                    -70                    -65
Asp Ser Cys Phe Gln Asp Leu Lys Glu Ser Arg Leu Val Glu Asp Thr
         -60                    -55                    -50
Phe Thr Ile Asp Glu Val Ser Glu Val Leu Asn Gly Leu Gln Ala Val
-45                    -40                    -35                    -30
Val His Ser Glu Val Glu Ser Glu Leu Ile Asn Thr Ala Tyr Thr Asn
                       -25                    -20                    -15
Val Leu Leu Leu Arg Gln Leu Phe Ala Gln Ala Glu Lys Trp Tyr Leu
              -10                    -5                     1
Lys Leu Gln Thr Asp Ile Ser Glu Leu Glu Asn Arg Glu Leu Leu Glu
         5                     10                    15
Gln Xaa Ala Glu Phe Glu Lys Ala Xaa Ile Thr Ser Ser Asn Lys Lys
20                    25                    30                    35
Pro Ile Leu Xaa Val Thr Xaa Pro Lys Leu Ala Pro Leu Asn Glu Gly
                   40                    45                    50
Gly Thr Ala Lys Leu Leu Asn Lys Val Ile Cys Ile Ile Leu Arg Asn
          55                    60                    65
Gly Lys Ser Leu Ile Leu Ser Cys His Cys Leu Gly Trp Arg Asn Lys
         70                    75                    80
Ser Gly Arg Phe Val Ser Gly Pro Leu Arg Ile Ile Ser Pro Leu Gln
         85                    90                    95
```


<210> 398
<211> 149
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -72..-1

<400> 398
```
Met Asn Leu Phe Ile Met Tyr Met Ala Gly Asn Thr Ile Ser Ile Phe
              -70                    -65                    -60
Pro Thr Met Met Val Cys Met Met Ala Trp Arg Pro Ile Gln Ala Leu
         -55                    -50                    -45
Met Ala Ile Ser Ala Thr Phe Lys Met Leu Glu Ser Ser Ser Gln Lys
-40                    -35                    -30                    -25
Phe Leu Gln Gly Leu Val Tyr Leu Ile Gly Asn Leu Met Gly Leu Ala
              -20                    -15                    -10
Leu Ala Val Tyr Lys Cys Gln Ser Met Gly Leu Leu Pro Thr His Ala
              -5                     1                     5
Ser Asp Trp Leu Ala Phe Ile Glu Pro Pro Glu Arg Met Glu Ser Val
         10                    15                    20
Val Glu Asp Cys Phe Cys Glu His Glu Lys Ala Ala Pro Gly Pro Tyr
25                    30                    35                    40
Val Phe Gly Ser Tyr Leu His Pro Ser Leu Ser Pro Val Ala Pro Gln
                   45                    50                    55
His Thr Leu Lys Leu Ile Thr Tyr Val Lys Lys Asn Gln Lys Thr Leu
              60                    65                    70
Phe Ser Met Val Gly
         75
```


<210> 399
<211> 73
<212> PRT
<213> Homo sapiens

```
<220>
<221> SIGNAL
<222> -20..-1

<400> 399
Met Thr Pro Leu Leu Thr Leu Ile Leu Val Val Leu Met Gly Leu Pro
-20                 -15                 -10                 -5
Leu Ala Gln Ala Leu Asp Cys His Val Cys Ala Tyr Asn Gly Asp Asn
                1               5               10
Cys Phe Asn Pro Met Arg Cys Pro Ala Met Val Ala Tyr Cys Met Thr
        15              20              25
Thr Arg Thr Tyr Tyr Thr Pro Thr Arg Met Lys Val Ser Lys Ser Cys
    30              35              40
Val Pro Arg Cys Phe Glu Xaa Cys Val
45                  50
```

```
<210> 400
<211> 86
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -20..-1

<400> 400
Met Asn Leu His Phe Pro Gln Trp Phe Val His Ser Ser Ala Leu Gly
-20                 -15                 -10                 -5
Leu Val Leu Ala Pro Pro Phe Ser Ser Pro Gly Thr Asp Pro Thr Phe
                1               5               10
Pro Cys Ile Tyr Cys Arg Leu Leu Asn Met Ile Met Thr Arg Leu Ala
        15              20              25
Phe Ser Phe Ile Thr Cys Leu Cys Pro Asn Leu Lys Glu Val Cys Leu
    30              35              40
Ile Leu Pro Glu Lys Asn Cys Asn Ser Arg His Ala Gly Phe Val Gly
45                  50              55                  60
Pro Xaa Lys Leu Arg Gln
                65
```

```
<210> 401
<211> 78
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 401
Met Cys Pro Val Phe Ser Lys Gln Leu Leu Ala Cys Gly Ser Leu Leu
    -20                 -15                 -10
Pro Gly Leu Trp Gln His Leu Thr Ala Asn His Trp Pro Pro Phe Ser
-5                  1               5               10
Xaa Phe Leu Cys Thr Val Cys Ser Gly Ser Ser Glu Gln Ile Ser Glu
        15              20              25
Tyr Thr Ala Ser Ala Thr Pro Pro Leu Cys Arg Ser Leu Asn Gln Glu
        30              35              40
Pro Phe Val Ser Arg Ala Ile Arg Pro Lys Tyr Ser Ile Thr
```

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9906549	A	11-02-1999	AU	8555098 A	22-02-1999
WO 9634981	A	07-11-1996	FR	2733765 A	08-11-1996
			FR	2733762 A	08-11-1996
			AU	5982996 A	21-11-1996
			CA	2220045 A	07-11-1996
			EP	0824598 A	25-02-1996
EP 0625572	A	23-11-1994	JP	6153953 A	03-06-1994
			WO	9408001 A	14-04-1994
			US	5597713 A	28-01-1997
WO 9707198	A	27-02-1997	US	5707829 A	13-01-1998
			AU	6712396 A	18-02-1997
			AU	6768596 A	12-03-1997
			CA	2227220 A	06-02-1997
			CA	2229208 A	27-02-1997
			EP	0839196 A	06-05-1998
			EP	0851875 A	08-07-1998
			WO	9704097 A	06-02-1997

FURTHER INFORMATION CONTINUED FROM   PCT/ISA/ 210

1. Claims: Invention 1: Claims 1-20, all partially.

   Nucleic acid comprising the sequence as in Seq.ID:40,
   complementary sequence or fragments, host cell containing
   said nucleic acid. Polypeptide as in Seq.ID:141, encoded by
   said polynucleotide, or fragments, method of making said
   polypeptide. Antibody specifically binding to said
   polypeptide.

2. Claims: Inventions 2-233: Claims 1-20, all partially,
              as far as applicable.

   Idem as subject 1 but limited to each of the DNA sequences
   as in Seq.ID:41-140, 242-377, and corresponding
   polypeptides, where invention 2 is limited to Seq.ID:41 and
   142, invention 3 is limited to Seq.ID:42 and 143, ...,
   invention 8 is limited to Seq.ID:47 and 148, invention 9 is
   limited to Seq.ID:48,49,110,149,150 and 211, invention 10 is
   limited to Seq.ID:50 and 151, ..., invention 32 is
   limited to Seq.ID:72 and 173, invention 33 is limited to
   Seq.ID:73,74,131,174,175 and 232, invention 34 is limited to
   Seq.ID:75 and 176, ..., invention 233 is limited to
   Seq.ID:377 and 513.

   For the sake of conciseness, the first subject matter is
   explicitly defined, the other subject matters are defined by
   analogy thereto.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB 98/02122

Box I   Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons.

1. [ ] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II   Observations where unity of Invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

> See additional sheet.

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos

> Invention 1, Claims 1-20 partially.

Remark on Protest   [ ] The additional search fees were accompanied by the applicant's protest

                               [ ] No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (1)) (July 1998)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LOCKHART D.J. ET AL.: "Expression monitoring by hybridization to high-density oligonucleotide arrays" BIO/TECHNOLOGY, no. 14, 14 December 1996 (1996-12-14), pages 1675-1680, XP002074420 abstract	18

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 34981 A (GENSET (FR); NICOLAEVNA MERENKOVA I.; DUMAS MILNE EDWARDS J.-B.G.) 7 November 1996 (1996-11-07) cited in the application abstract	
A	EP 0 625 572 A (KANAGAWA ACAD OF SCIENCE AND TECHNOL FOUNDATION (JP); KATO S; SEKINE S) 23 November 1994 (1994-11-23) cited in the application abstract	
A	CARNINCI P. ET AL.: "High-efficiency full-length cDNA cloning by biotinylated CAP trapper" GENOMICS, vol. 37, no. 3, 1 November 1996 (1996-11-01), pages 327-336, XP002081729 cited in the application abstract	
A	KATO S. ET AL.: "Construction of a human full-length cDNA bank" GENE, vol. 150, 1994, pages 243-250, XP002081364 cited in the application abstract	
A	WO 97 07198 A (GENETICS INSTITUTE INC (US); JACOBS K; MCCOY JM; KELLEHER K; CARLIN M) 27 February 1997 (1997-02-27)	
A	TASHIRO K. ET AL.: "Signal sequence trap: a cloning strategy for secreted proteins and type I membrane proteins" SCIENCE, vol. 261, 30 July 1993 (1993-07-30), pages 600-603, XP000673204 abstract	
A	YOKOYAMA-KOBAYASHI M. ET AL.: "A signal sequence detection system using secreted protease activity as an indicator" GENE, vol. 163, 1995, pages 193-196, XP002053953 abstract	
A	HEIJNE VON G.: "A new method for predicting signal sequence cleavage sites" NUCLEIC ACIDS RESEARCH, vol. 14, no. 11, 1986, pages 4683-4690, XP002053954 cited in the application abstract	

-/--

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6   C12N15/12   C07K14/47   C07K16/18   C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6   C12N   C07K   C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E,L	WO 99 06549 A (GENSET (FR); DUMAS MILNE EDWARDS J.-B.; DUCLERT A.; LACROIX B.) 11 February 1999 (1999-02-11) L: Priority abstract page 6 - page 12 page 129 - page 133; claims Seq.ID:251 page 213 - page 214 Seq.ID:484 page 366 - page 367 ---	1-20
X	Database EMBL, entry HS695112 Accession number R50695 24 May 1995 95% identity with Seq.ID:40 nt.1-384 XP002097725 the whole document ---	2,5,8

-/--

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 March 1999	27. 07. 99
Name and mailing address of the ISA European Patent Office, P B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk . Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Macchia, G

Form PCT/ISA/210 (second sheet) (July 1992)

PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/47, 16/18, C12Q 1/68	**A3**	**(11) International Publication Number:** WO 99/31236
		(43) International Publication Date: 24 June 1999 (24.06.99)

(21) International Application Number: PCT/IB98/02122

(22) International Filing Date: 17 December 1998 (17.12.98)

(30) Priority Data:

60/069,957	17 December 1997 (17.12.97)	US
60/074,121	9 February 1998 (09.02.98)	US
60/081,563	13 April 1998 (13.04.98)	US
60/096,116	10 August 1998 (10.08.98)	US

(71) Applicant *(for all designated States except US)*: GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BOUGUELERET, Lydie [FR/FR]; 108, avenue Victor Hugo, F-92170 Vanves (FR). DUCLERT, Aymeric [FR/FR]; 6 ter, rue Victorine, F-94100 Saint-Maur (FR). DUMAS MILNE EDWARDS, Jean-Baptiste [FR/FR]; 8, rue Grégoire de Tours, F-75006 Paris (FR).

(74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
10 September 1999 (10.09.99)

---

(54) Title: EXTENDED cDNAs FOR SECRETED PROTEINS

(57) Abstract

The sequences of extended cDNAs encoding secreted proteins are disclosed. The extended cDNAs can be used to express secreted proteins or portions thereof or to obtain antibodies capable of specifically binding to the secreted proteins. The extended cDNAs may also be used in diagnostic, forensic, gene therapy, and chromosome mapping procedures. The extended cDNAs may also be used to design expression vectors and secretion vectors.

45 50 55

<210> 402
<211> 65
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -28..-1

<400> 402
Met Gly Lys Gly His Gln Arg Pro Trp Trp Lys Val Leu Pro Leu Ser
        -25   -20   -15
Cys Phe Leu Val Ala Leu Ile Ile Trp Cys Tyr Leu Arg Glu Glu Ser
    -10   -5   1
Glu Ala Asp Gln Trp Leu Arg Gln Val Trp Gly Glu Val Pro Glu Pro
5   10   15   20
Ser Asp Arg Ser Glu Glu Pro Glu Thr Pro Ala Ala Tyr Arg Ala Arg
                25   30   35
Thr

<210> 403
<211> 211
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -27..-1

<400> 403
Met Leu Leu Leu Ser Ile Thr Thr Ala Tyr Thr Gly Leu Glu Leu Thr
        -25   -20   -15
Phe Phe Ser Gly Val Tyr Gly Thr Cys Ile Gly Ala Thr Asn Lys Phe
    -10   -5   1   5
Gly Ala Glu Glu Xaa Ser Leu Ile Gly Leu Ser Gly Ile Phe Ile Gly
                10   15   20
Ile Gly Glu Ile Leu Gly Gly Ser Leu Phe Gly Leu Leu Ser Lys Asn
            25   30   35
Asn Arg Phe Gly Arg Asn Pro Val Leu Leu Gly Ile Leu Val His
        40   45   50
Phe Ile Ala Phe Tyr Leu Ile Phe Leu Asn Met Pro Gly Asp Ala Pro
    55   60   65
Ile Ala Pro Val Lys Gly Thr Asp Ser Ser Ala Tyr Ile Lys Ser Ser
70   75   80   85
Lys Xaa Phe Ala Ile Leu Cys Xaa Phe Leu Xaa Gly Leu Gly Asn Ser
            90   95   100
Cys Phe Asn Thr Xaa Leu Leu Xaa Ile Xaa Gly Phe Leu Tyr Ser Glu
            105   110   115
Xaa Ser Ala Pro Xaa Phe Ala Ile Phe Asn Phe Val Gln Ser Ile Cys
        120   125   130
Ala Ala Val Ala Phe Phe Tyr Ser Asn Tyr Leu Leu Leu His Trp Gln
        135   140   145
Leu Leu Val Met Val Ile Phe Gly Phe Xaa Gly Thr Ile Ser Phe Phe
150   155   160   165
Thr Val Glu Trp Glu Xaa Ala Ala Phe Val Xaa Arg Gly Ser Asp Tyr
                170   175   180

Arg Ser Ile

<210> 404
<211> 123
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -80..-1

<400> 404
Met Ser Thr Trp Tyr Leu Ala Leu Asn Lys Ser Tyr Lys Asn Lys Asp
-80   -75   -70   -65
Ser Val Arg Ile Tyr Leu Ser Leu Cys Thr Val Ser Ile Lys Phe Thr
            -60   -55   -50
Tyr Phe His Asp Ile Gln Thr Asn Cys Leu Thr Thr Trp Lys His Ser
            -45   -40   -35
Arg Cys Arg Phe Tyr Trp Ala Phe Gly Gly Ser Ile Leu Gln His Ser
        -30   -25   -20
Val Asp Pro Leu Val Leu Phe Leu Ser Leu Ala Leu Leu Val Thr Pro
    -15   -10   -5
Thr Ser Thr Pro Ser Ala Lys Ile Gln Ser Leu Gln Ile Asp Leu Pro
1   5   10   15
Gly Gly Trp Arg Leu Ala Thr Asp Arg Ile Phe Thr Leu Ser Pro Val
            20   25   30
Pro Met Asp Xaa Pro Leu Ile Leu His Gln Leu
        35   40

<210> 405
<211> 86
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -26..-1

<400> 405
Met Glu Lys Ser Trp Met Leu Trp Asn Phe Val Glu Arg Trp Leu Ile
        -25   -20   -15
Ala Leu Ala Ser Trp Ser Trp Ala Leu Cys Arg Ile Ser Leu Leu Pro
-10   -5   1   5
Leu Ile Val Thr Phe His Leu Tyr Gly Gly Ile Ile Leu Leu Leu Leu
            10   15   20
Ile Phe Ile Ser Ile Xaa Gly Ile Leu Tyr Lys Phe Xaa Asp Val Leu
        25   30   35
Leu Tyr Phe Pro Xaa Gln Xaa Ser Ser Ser Arg Leu Tyr Asp Ser His
    40   45   50
Ala His Trp Xaa Ser Xaa
55   60

<210> 406
<211> 162
<212> PRT
<213> Homo sapiens

```
<220>
<221> SIGNAL
<222> -31..-1

<400> 406
Met Ala Ala Ala Trp Pro Ser Gly Pro Xaa Ala Pro Glu Ala Val Thr
    -30                     -25                 -20
Ala Arg Leu Val Gly Val Leu Trp Phe Val Ser Val Thr Thr Gly Pro
-15                     -10                 -5                  1
Trp Gly Ala Val Ala Thr Ser Ala Gly Gly Glu Glu Ser Leu Lys Cys
            5                   10                  15
Glu Asp Leu Lys Val Gly Gln Tyr Ile Cys Lys Asp Pro Lys Ile Asn
        20                  25                  30
Asp Ala Thr Gln Glu Pro Val Asn Cys Thr Asn Tyr Thr Ala His Val
    35                  40                  45
Ser Cys Phe Pro Ala Pro Asn Ile Thr Cys Lys Asp Ser Ser Gly Asn
50                  55                  60                      65
Glu Thr His Phe Thr Gly Asn Glu Val Gly Phe Phe Lys Pro Ile Ser
                70                  75                  80
Cys Arg Asn Val Asn Gly Tyr Ser Tyr Asn Glu Gln Ser His Val Ser
            85                  90                  95
Phe Ser Trp Met Val Gly Ser Arg Ser Ile Leu Pro Trp Ile Pro Cys
            100                 105                 110
Phe Gly Phe Val Lys Xaa Xaa His Cys Arg Val Xaa Trp Asn Trp Glu
    115                 120                 125
Pro Asn
130


<210> 407
<211> 98
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -37..-1

<400> 407
Met Ala Ser Leu Leu Cys Cys Gly Pro Lys Leu Ala Ala Cys Gly Ile
        -35                 -30                 -25
Val Leu Ser Ala Trp Gly Val Ile Met Leu Ile Met Leu Gly Ile Phe
    -20                 -15                 -10
Phe Asn Val His Ser Ala Val Leu Ile Glu Asp Val Pro Phe Thr Glu
-5                  1               5                   10
Lys Asp Phe Glu Asn Gly Pro Gln Asn Ile Tyr Asn Leu Tyr Xaa Gln
            15                  20                  25
Xaa Ser Tyr Asn Cys Phe Ile Ala Ala Gly Leu Tyr Leu Leu Leu Gly
        30                  35                  40
Gly Phe Ser Phe Cys Gln Xaa Arg Leu Asn Lys Arg Lys Glu Tyr Met
    45                  50                  55
Val Arg
60


<210> 408
<211> 70
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -15..-1

<400> 408
Met Arg Phe Leu Pro Cys Cys Leu Leu Trp Ser Val Phe Asn Pro Glu
-15                 -10              -5                       1
Ser Leu Asn Cys His Tyr Phe Xaa Xaa Glu Xaa Cys Ile Phe Xaa Ser
            5               10                  15
Leu Gln Tyr Tyr Glu Ile Ser Leu Gln Glu Lys Leu Leu Gly Phe Leu
        20              25              30
Trp Leu Cys Phe Leu Ser Tyr Phe Phe Arg Ala Val Tyr Phe Leu Ile
    35              40                  45
Asp Phe Ser Ser Phe Thr
50              55



<210> 409
<211> 60
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -45..-1

<400> 409
Met His Ser Leu Phe Ile Ala Ser Leu Lys Val Leu Phe Tyr Tyr Ser
-45                 -40             -35                      -30
Phe Ser Phe Arg Phe Asn Trp Phe Asp Cys Leu Leu His Asn Leu Gly
            -25             -20                 -15
Glu Asn Phe Leu Ser Leu Leu Ser Lys Ser Cys Ser Ala Asp Pro Ser
            -10             -5                  1
Gly Ser Thr Phe Met Arg Asp Ile Glu Thr Asn Lys
    5               10                  15



<210> 410
<211> 39
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -22..-1

<400> 410
Met Pro Glu Ala Val Glu Gln Ser Ala His Leu Phe Val Thr Trp Ser
        -20             -15             -10
Ser Gln Arg Ala Leu Ser His Pro Ala Pro Phe Leu Thr Xaa Xaa Lys
    -5                  1               5                   10
Asn Pro Phe Leu Trp Lys Leu
            15



<210> 411
<211> 51
<212> PRT
```

```
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -23..-1

<400> 411
Met Ala Phe Gln Ser Leu Leu Glu Met Lys Phe Phe Leu Cys Ala Ala
            -20                 -15                 -10
Phe Pro Leu Gly Ala Gly Val Lys Met Phe His Tyr Leu Gly Pro Gly
         -5                  1                  5
Lys Pro Leu Xaa Gln Ala Ser Pro Ser Pro His Pro His Arg Xaa Arg
10                  15                  20                  25
Ile Trp Pro



<210> 412
<211> 95
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -48..-1

<400> 412
Met Ala Ser Ser His Trp Asn Glu Thr Thr Thr Ser Val Tyr Gln Tyr
            -45                 -40                 -35
Leu Gly Phe Gln Val Gln Lys Ile Tyr Pro Phe His Asp Asn Trp Asn
         -30                 -25                 -20
Thr Ala Cys Phe Val Ile Leu Leu Leu Phe Ile Phe Thr Val Val Ser
         -15                 -10                  -5
Leu Val Val Leu Ala Phe Leu Tyr Glu Val Leu Xaa Xaa Cys Cys Cys
1                  5                  10                  15
Val Lys Asn Lys Thr Val Lys Asp Leu Lys Ser Glu Pro Asn Pro Leu
            20                  25                  30
Xaa Xaa Met Met Asp Asn Ile Arg Lys Arg Glu Thr Glu Val Val
         35                  40                  45



<210> 413
<211> 60
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -32..-1

<400> 413
Met Asp Glu Tyr Ser Trp Trp Cys His Val Leu Glu Val Val Lys Gly
            -30                 -25                 -20
Gln Met Phe Thr Phe Ile Asn Ile Thr Leu Trp Leu Gly Ser Leu Cys
         -15                 -10                  -5
Gln Arg Phe Phe Tyr Ala Ser Gly Thr Tyr Phe Leu Ile Tyr Ile Ser
1                  5                  10                  15
Thr Val Thr Pro Ser Trp Arg Leu Cys Leu Val Ser
            20                  25
```

```
<210> 414
<211> 170
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -79..-1

<400> 414
Met Glu Asp Pro Asn Pro Glu Glu Asn Met Lys Gln Gln Asp Ser Pro
            -75             -70             -65
Lys Glu Arg Ser Pro Gln Ser Pro Gly Gly Asn Ile Cys His Leu Gly
            -60             -55             -50
Ala Pro Lys Cys Thr Arg Cys Leu Ile Thr Phe Ala Asp Ser Lys Phe
            -45             -40             -35
Gln Glu Arg His Met Lys Arg Glu His Pro Ala Asp Phe Val Ala Gln
        -30             -25             -20
Lys Leu Gln Gly Val Leu Phe Ile Cys Phe Thr Cys Ala Arg Ser Phe
-15             -10             -5              1
Pro Ser Ser Lys Ala Xaa Xaa Thr His Gln Arg Ser His Gly Pro Xaa
            5               10              15
Ala Lys Pro Thr Leu Pro Val Ala Thr Thr Thr Ala Gln Pro Thr Phe
        20              25              30
Pro Cys Pro Asp Cys Gly Lys Thr Phe Gly Gln Ala Val Ser Leu Xaa
    35              40              45
Arg His Xaa Gln Xaa His Glu Val Arg Ala Pro Pro Gly Thr Phe Ala
50              55              60              65
Cys Thr Xaa Cys Gly Gln Asp Phe Ala Gln Glu Xaa Gly Leu His Gln
            70              75              80
His Tyr Ile Arg His Ala Arg Gly Gly Leu
            85              90


<210> 415
<211> 190
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -82..-1

<400> 415
Met Tyr Val Trp Pro Cys Ala Val Val Leu Ala Gln Tyr Leu Trp Phe
        -80             -75             -70
His Arg Arg Ser Leu Pro Gly Lys Ala Ile Leu Glu Ile Gly Ala Gly
        -65             -60             -55
Val Ser Leu Pro Gly Ile Leu Ala Ala Lys Cys Gly Ala Glu Val Ile
-50             -45             -40             -35
Leu Ser Asp Ser Ser Glu Leu Pro His Cys Leu Glu Val Cys Arg Gln
            -30             -25             -20
Ser Cys Gln Met Asn Asn Leu Pro His Leu Gln Val Val Gly Leu Thr
            -15             -10             -5
Trp Gly His Ile Ser Trp Asp Leu Leu Ala Leu Pro Pro Gln Asp Ile
        1               5               10
Ile Leu Ala Ser Asp Val Phe Phe Glu Pro Glu Xaa Phe Glu Asp Ile
15              20              25              30
Leu Ala Thr Ile Tyr Phe Leu Met His Lys Asn Pro Lys Val Gln Leu
            35              40              45
```

```
Trp Ser Thr Tyr Gln Val Arg Xaa Ala Asp Trp Ser Leu Glu Ala Leu
         50                  55              60
Leu Tyr Lys Trp Asp Met Lys Cys Val His Ile Pro Leu Glu Ser Phe
         65                  70              75
Asp Ala Asp Lys Glu Xaa Ile Ala Glu Ser Thr Leu Pro Gly Arg His
     80                  85              90
Thr Val Glu Met Leu Val Ile Ser Phe Ala Lys Asp Ser Leu
95                 100             105
```

```
<210> 416
<211> 114
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -60..-1

<400> 416
Met Met Ala Ala Val Pro Pro Gly Leu Glu Pro Trp Asn Arg Val Arg
-60             -55             -50                 -45
Ile Pro Lys Ala Gly Asn Arg Ser Ala Val Thr Val Gln Asn Pro Gly
            -40             -35                 -30
Ala Ala Leu Asp Leu Cys Ile Ala Ala Val Ile Lys Glu Cys His Leu
            -25             -20                 -15
Val Ile Leu Ser Leu Lys Ser Gln Thr Leu Asp Ala Glu Thr Asp Val
            -10             -5              1
Leu Cys Ala Val Leu Tyr Ser Asn His Asn Arg Met Gly Arg His Lys
5               10              15                  20
Pro His Leu Ala Leu Lys Gln Val Glu Gln Cys Leu Lys Arg Leu Lys
            25              30                  35
Asn Met Asn Leu Glu Gly Ser Ile Gln Asp Leu Phe Glu Leu Phe Ser
            40              45                  50
Ser Lys
```

```
<210> 417
<211> 161
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -108..-1

<400> 417
Met Thr Ser Gly Gln Ala Arg Ala Ser Xaa Gln Ser Pro Gln Ala Leu
            -105            -100            -95
Glu Asp Ser Gly Pro Val Asn Ile Ser Val Ser Ile Thr Leu Thr Leu
         -90             -85             -80
Asp Pro Leu Lys Pro Phe Gly Gly Tyr Ser Arg Asn Val Thr His Leu
     -75             -70             -65
Tyr Ser Thr Ile Leu Gly His Gln Ile Gly Leu Ser Gly Arg Glu Ala
-60             -55             -50                 -45
His Glu Glu Ile Asn Ile Thr Phe Thr Leu Pro Thr Ala Trp Ser Ser
            -40             -35             -30
Asp Asp Cys Ala Leu His Gly His Cys Glu Gln Val Val Phe Thr Ala
            -25             -20                 -15
Cys Met Thr Leu Thr Ala Ser Pro Gly Val Phe Pro Ser Leu Tyr Ser
```

```
                -10                      -5                       1
His Arg Thr Val Phe Leu Thr Arg Thr Ala Thr Pro Arg Ser Gly Thr
 5                   10                       15                     20
Arg Ser Ser Gln Leu Pro Glu Met Pro Thr Gln Asn Thr Pro Lys Ile
                   25                       30                     35
Thr Ile Leu Ser Gly Val Ile Arg Gly Pro Leu Glu Lys Ser Ile Met
                   40                       45                     50
Leu
```

```
<210> 418
<211> 67
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 418
Met Leu Gly Gly Asp His Arg Ala Leu Leu Leu Lys Ile Trp Leu Leu
    -20                     -15                  -10
Gln Arg Pro Glu Ser Gln Glu Gly Leu Leu Pro Gly Arg Leu Val Val
-5                      1                   5                    10
Met Glu Arg Arg Val Lys Asn Asp Leu Met Ser Phe Leu Ser Thr Val
               15                       20                  25
Leu Leu Ser Phe His Ser Ser Asn Ala Arg Val Ser His Cys Glu Pro
          30                      35                  40
Leu Arg Met
    45
```

```
<210> 419
<211> 332
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -32..-1

<400> 419
Met Ile Xaa Leu Arg Asp Thr Ala Ala Ser Leu Arg Leu Glu Arg Asp
         -30                     -25                  -20
Thr Arg Gln Leu Pro Leu Leu Thr Ser Ala Leu His Gly Leu Gln Gln
    -15                 -10                     -5
Gln His Pro Ala Phe Ser Gly Val Ala Arg Leu Ala Lys Arg Trp Val
1                   5                       10                     15
Arg Ala Gln Leu Leu Gly Glu Gly Phe Ala Asp Glu Ser Leu Asp Leu
               20                      25                     30
Val Ala Ala Ala Leu Phe Leu His Pro Glu Pro Phe Thr Pro Pro Ser
          35                      40                     45
Ser Pro Gln Val Gly Phe Leu Arg Phe Leu Phe Leu Val Ser Thr Phe
    50                      55                     60
Asp Trp Lys Asn Asn Pro Leu Phe Val Asn Leu Asn Asn Glu Leu Thr
65                      70                      75                     80
Val Glu Glu Gln Val Glu Ile Arg Ser Gly Phe Leu Ala Ala Arg Ala
               85                      90                     95
Gln Leu Pro Val Met Val Ile Val Thr Pro Gln Xaa Arg Lys Asn Ser
              100                     105                    110
```

```
Val Trp Thr Gln Asp Gly Pro Ser Ala Gln Ile Leu Gln Gln Leu Val
        115                  120              125
Val Leu Ala Ala Glu Xaa Leu Pro Met Leu Xaa Xaa Gln Leu Met Asp
    130                  135              140
Pro Arg Gly Pro Gly Asp Ile Arg Thr Xaa Phe Arg Pro Pro Leu Asp
145                  150                  155              160
Ile Tyr Asp Val Leu Ile Arg Leu Ser Pro Arg His Ile Pro Arg His
            165                  170                  175
Arg Gln Ala Val Asp Ser Pro Ala Ala Ser Phe Cys Arg Gly Leu Leu
        180                  185                  190
Ser Gln Pro Gly Pro Ser Ser Leu Met Pro Val Leu Gly Xaa Asp Pro
        195                  200                  205
Pro Gln Leu Tyr Leu Thr Gln Leu Xaa Glu Ala Phe Gly Asp Leu Ala
    210                  215                  220
Leu Phe Phe Tyr Asp Gln His Gly Gly Glu Val Ile Gly Val Leu Trp
225                  230                  235              240
Lys Pro Thr Ser Phe Gln Pro Gln Pro Phe Lys Ala Ser Ser Thr Lys
            245                  250                  255
Gly Arg Met Val Met Ser Arg Gly Gly Glu Leu Val Met Val Pro Asn
        260                  265                  270
Val Glu Ala Ile Leu Glu Asp Phe Ala Val Leu Gly Glu Gly Leu Val
    275                  280                  285
Gln Thr Val Glu Ala Arg Ser Glu Arg Trp Thr Val
    290                  295                  300


<210> 420
<211> 65
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -19..-1

<400> 420
Met Gly Gly Ile Trp Asn Ala Leu Ser Met Ser Ser Phe Ser Phe His
            -15                  -10                  -5
Ser Ser Ser Cys Ser Ala Leu Ser Ala Lys Ser Leu Leu Ser Arg His
        1                    5                    10
His Ile Leu Gln Gln Phe Leu Val Arg Lys Ser Val Pro Leu Glu Asn
    15                   20                   25
Ala Ser Leu Pro Phe Pro His Leu Gly Ser Ser Leu Phe Lys Ile Val
30                   35                   40                   45
Gly


<210> 421
<211> 57
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -30..-1

<400> 421
Met Pro Thr Gly Lys Gln Leu Ala Asp Ile Gly Tyr Lys Thr Phe Ser
-30                  -25                  -20                  -15
Thr Ser Met Met Leu Leu Thr Val Tyr Gly Gly Tyr Leu Cys Ser Val
```

```
                              -10                    -5                       1
          Arg Val Tyr His Tyr Phe Gln Trp Arg Arg Ala Gln Arg Gln Ala Ala
                   5                      10                   15
          Glu Glu Gln Lys Xaa Ser Gly Ile Met
                  20                   25



          <210> 422
          <211> 85
          <212> PRT
          <213> Homo sapiens

          <220>
          <221> SIGNAL
          <222> -17..-1

          <400> 422
          Met Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val
                  -15                   -10                  -5
          Gly Phe Pro Val Ser Gln Asp Gln Glu Arg Glu Lys Arg Ser Ile Ser
              1               5                   10                   15
          Asp Ser Asp Glu Leu Ala Ser Gly Xaa Phe Val Phe Pro Tyr Pro Tyr
                          20                  25                   30
          Pro Phe Arg Pro Leu Pro Pro Ile Pro Phe Pro Arg Phe Pro Trp Phe
                      35                  40                   45
          Arg Arg Asn Phe Pro Ile Pro Ile Pro Glu Ser Ala Pro Thr Thr Pro
                      50                  55                   60
          Leu Pro Ser Glu Lys
              65



          <210> 423
          <211> 85
          <212> PRT
          <213> Homo sapiens

          <220>
          <221> SIGNAL
          <222> -17..-1

          <400> 423
          Met Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val
                  -15                   -10                  -5
          Gly Phe Pro Val Ser Gln Asp Xaa Glu Arg Glu Lys Arg Ser Ile Ser
              1               5                   10                   15
          Asp Ser Asp Glu Leu Ala Ser Gly Phe Phe Val Phe Pro Tyr Pro Tyr
                          20                  25                   30
          Pro Phe Arg Pro Leu Pro Pro Ile Pro Phe Pro Arg Phe Pro Trp Phe
                      35                  40                   45
          Arg Arg Asn Phe Pro Ile Pro Ile Pro Glu Ser Ala Pro Thr Thr Pro
                      50                  55                   60
          Leu Pro Ser Glu Lys
              65



          <210> 424
          <211> 69
          <212> PRT
          <213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -29..-1

<400> 424
Met Thr Cys Arg Gly Ser Cys Ser Tyr Ala Thr Arg Arg Ser Pro Ser
              -25                 -20                 -15
Glu Leu Ser Leu Leu Pro Ser Ser Leu Trp Val Leu Ala Thr Ser Ser
              -10                 -5                   1
Pro Thr Ile Thr Ile Ala Leu Ala Met Ala Ala Gly Asn Leu Cys Pro
         5                   10                  15
Leu Pro Ser Ser Xaa Arg Xaa Lys Arg Arg Trp Cys Gln Ala Xaa Gln
20                  25                  30                  35
Gln Xaa Ala Leu Leu
              40




<210> 425
<211> 122
<212> PRT
<213> Homo sapiens


<220>
<221> SIGNAL
<222> -56..-1

<400> 425
Met Val Pro Trp Pro Arg Gly Lys Val Lys Thr Ala Pro Ile Pro Ile
    -55                 -50                 -45
Ser Arg Phe Pro Phe Leu Pro Thr His Asp Pro Pro Thr Pro Ala His
-40                 -35                 -30                 -25
Trp Ser Pro Ala Ser His Gln Gln Phe Lys His Xaa Ser Pro Leu Leu
              -20                 -15                 -10
Thr Leu Ala Leu Leu Gly Gln Cys Ser Leu Phe Xaa Asn Leu Arg Lys
              -5                   1                   5
Lys Leu Ala Gly Gln Lys Ala Lys Lys Leu Pro Ser Phe Ser Ser Leu
    10                  15                  20
Pro Leu Thr Leu Trp Pro Leu Thr Pro Gln Phe Ala Glu Leu Thr Thr
25                  30                  35                  40
Val Ala Gln Lys Lys Leu Arg Trp Ser Gly Thr Leu Gly Trp Gly Pro
              45                  50                  55
Val Pro Ser Trp Val Gln Phe Phe Leu Gly
              60                  65




<210> 426
<211> 41
<212> PRT
<213> Homo sapiens


<220>
<221> SIGNAL
<222> -30..-1

<400> 426
Met Ala Cys Glu Thr His Gly Val Leu Val Pro Ala His Leu Ser Gly
-30                 -25                 -20                 -15
Leu Ile Thr Cys Leu Leu Ala Phe Trp Val Pro Ala Ser Cys Ile Gln
              -10                 -5                   1
```

```
Arg Cys Ser Gly Ser Pro Leu Pro Leu
        5                   10
```


<210> 427
<211> 50
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -36..-1

<400> 427
```
Met Ala Pro His Thr Ala Ser Phe Gly Val Cys Pro Leu Leu Ser Val
    -35                 -30                 -25
Thr Arg Val Val Ala Thr Glu His Trp Leu Phe Leu Ala Ser Leu Ser
    -20                 -15                 -10                 -5
Gly Ile Lys Thr Tyr Gln Ser Tyr Ile Ser Val Phe Cys Lys Val Thr
              1                   5                   10
Leu Ile
```


<210> 428
<211> 136
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -18..-1

<400> 428
```
Met Asp Ser Leu Arg Lys Met Leu Ile Ser Val Ala Met Leu Gly Ala
              -15                 -10                 -5
Xaa Ala Gly Val Gly Tyr Ala Leu Leu Val Ile Val Thr Pro Gly Glu
    1                   5                   10
Arg Arg Lys Gln Glu Met Leu Lys Glu Met Pro Leu Gln Asp Pro Arg
15                  20                  25                  30
Ser Arg Glu Glu Ala Ala Arg Thr Gln Gln Leu Leu Leu Ala Thr Leu
              35                  40                  45
Gln Glu Ala Ala Thr Thr Gln Glu Asn Val Ala Trp Arg Lys Asn Trp
              50                  55                  60
Met Val Gly Gly Glu Gly Gly Ala Thr Gly Xaa His Arg Glu Thr Gly
          65                  70                  75
Leu Ala Ser Val Gly Ala Gly Pro Trp Leu Gly Arg Arg Asn Pro Arg
    80                  85                  90
Gln Leu Ser Pro Ser Trp Ala Xaa Arg Lys Ile Arg Xaa Glu Asn Xaa
95                  100                 105                 110
Met Pro Gly Leu Ser Gly Val Leu
              115
```


<210> 429
<211> 194
<212> PRT
<213> Homo sapiens

<220>

<221> SIGNAL
<222> -65..-1

<400> 429
Met Gln Asp Ala Pro Leu Ser Cys Leu Ser Pro Thr Lys Trp Ser Ser
-65   -60   -55   -50
Val Ser Ser Ala Asp Ser Thr Glu Lys Ser Ala Ser Ala Ala Gly Thr
            -45   -40   -35
Arg Asn Leu Pro Phe Gln Phe Cys Leu Arg Gln Ala Leu Arg Met Lys
        -30   -25   -20
Ala Ala Gly Ile Leu Thr Leu Ile Gly Cys Leu Val Thr Gly Val Glu
        -15   -10   -5
Ser Lys Ile Tyr Thr Arg Cys Lys Leu Ala Lys Ile Phe Ser Arg Ala
1   5   10   15
Gly Leu Asp Asn Xaa Arg Gly Phe Ser Leu Gly Asn Trp Ile Cys Met
            20   25   30
Ala Tyr Tyr Glu Ser Gly Tyr Asn Thr Thr Ala Gln Thr Val Leu Asp
        35   40   45
Asp Gly Ser Ile Asp Tyr Gly Ile Phe Gln Ile Asn Ser Phe Ala Trp
        50   55   60
Cys Arg Arg Gly Lys Leu Lys Glu Asn Asn His Cys His Val Ala Cys
    65   70   75
Ser Ala Leu Xaa Thr Asp Asp Leu Thr Asp Ala Ile Ile Cys Ala Xaa
80   85   90   95
Lys Ile Val Lys Glu Thr Gln Gly Met Asn Tyr Trp Gln Gly Trp Lys
            100   105   110
Lys His Cys Glu Gly Arg Asp Leu Ser Xaa Trp Lys Lys Gly Cys Glu
            115   120   125
Val Ser

<210> 430
<211> 141
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -69..-1

<400> 430
Met Thr Ser Gln Pro Val Pro Asn Glu Thr Ile Ile Val Leu Pro Ser
            -65   -60   -55
Asn Val Ile Asn Phe Ser Gln Ala Glu Lys Pro Glu Pro Thr Asn Gln
            -50   -45   -40
Gly Gln Asp Ser Leu Lys Lys His Leu His Ala Glu Ile Lys Val Ile
        -35   -30   -25
Gly Thr Ile Gln Ile Leu Cys Gly Met Met Val Leu Ser Leu Gly Ile
    -20   -15   -10
Ile Leu Ala Ser Ala Ser Phe Ser Pro Asn Phe Thr Gln Val Thr Ser
-5   1   5   10
Thr Leu Leu Asn Ser Ala Tyr Pro Phe Ile Gly Pro Phe Phe Val Xaa
            15   20   25
Lys Xaa Ser Glu Glu Gly Arg Met Gly Gln Xaa Gly Glu Glu Xaa Xaa
        30   35   40
Asn Ser Leu Asn Phe Pro Xaa Ala Ser Leu Leu Xaa Leu Ile Cys Gln
    45   50   55
Xaa Gln Gly Phe Asn Gly Glu Ser Cys Ser Pro Val Gly
60   65   70

```
<210> 431
<211> 248
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -69..-1

<400> 431
Met Thr Ser Gln Pro Val Pro Asn Glu Thr Ile Ile Val Leu Pro Ser
              -65                 -60                 -55
Asn Val Ile Asn Phe Ser Gln Ala Glu Lys Pro Glu Pro Thr Asn Gln
              -50                 -45                 -40
Gly Gln Asp Ser Leu Lys Lys His Leu His Ala Glu Xaa Lys Val Ile
              -35                 -30                 -25
Gly Thr Ile Gln Ile Leu Cys Gly Met Met Val Leu Ser Leu Gly Ile
              -20                 -15                 -10
Ile Leu Ala Ser Ala Ser Phe Ser Pro Asn Phe Thr Gln Val Thr Ser
-5                  1                 5                   10
Thr Leu Leu Asn Ser Ala Tyr Pro Phe Ile Gly Pro Phe Phe Phe Ile
              15                  20                  25
Ile Ser Gly Ser Leu Ser Ile Ala Thr Lys Lys Arg Leu Thr Asn Leu
              30                  35                  40
Leu Val His Thr Thr Leu Val Gly Ser Ile Leu Ser Ala Leu Ser Ala
              45                  50                  55
Leu Val Gly Phe Ile Xaa Leu Ser Val Lys Gln Ala Thr Leu Asn Pro
60                  65                  70                  75
Ala Ser Leu Xaa Cys Glu Leu Xaa Lys Asn Asn Ile Pro Thr Xaa Xaa
              80                  85                  90
Tyr Val Xaa Tyr Phe Tyr His Asp Ser Leu Tyr Thr Thr Asp Xaa Tyr
              95                  100                 105
Thr Ala Lys Ala Xaa Leu Ala Gly Thr Leu Ser Leu Met Leu Ile Cys
              110                 115                 120
Thr Leu Leu Glu Phe Cys Xaa Xaa Val Leu Thr Ala Val Leu Arg Trp
              125                 130                 135
Lys Gln Ala Tyr Ser Asp Phe Pro Gly Ser Val Leu Phe Leu Pro Xaa
140                 145                 150                 155
Ser Tyr Ile Gly Asn Ser Gly Met Ser Ser Lys Met Thr His Asp Cys
              160                 165                 170
Gly Tyr Glu Glu Leu Leu Thr Ser
              175
```

```
<210> 432
<211> 49
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -36..-1

<400> 432
Met Gln Val Pro His Leu Arg Val Trp Thr Gln Val Xaa Asp Thr Phe
              -35                 -30                 -25
Ile Gly Tyr Arg Asn Leu Gly Phe Thr Ser Met Cys Ile Leu Phe His
-20                 -15                 -10                 -5
Cys Leu Leu Ser Phe Gln Val Phe Lys Lys Lys Arg Lys Leu Xaa Leu
              1                   5                   10
```

Phe

<210> 433
<211> 86
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -14..-1

<400> 433
Met Val Ala Leu Asn Leu Ile Leu Val Pro Cys Cys Ala Ala Trp Cys
             -10   -5   1
Asp Pro Arg Arg Ile His Ser Gln Asp Asp Val Leu Arg Ser Ser Ala
         5   10   15
Ala Asp Thr Gly Ser Ala Met Gln Arg Arg Glu Ala Trp Ala Gly Trp
         20   25   30
Arg Arg Ser Gln Pro Phe Ser Val Gly Leu Pro Ser Ala Glu Arg Leu
35   40   45   50
Glu Asn Gln Pro Gly Lys Leu Ser Trp Arg Ser Leu Val Gly Glu Gly
                55   60   65
His Arg Ile Cys Asp Leu
             70

<210> 434
<211> 144
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -58..-1

<400> 434
Met Thr Arg Leu Cys Leu Pro Arg Pro Glu Ala Arg Glu Asp Pro Ile
             -55   -50   -45
Pro Val Pro Pro Arg Gly Leu Gly Ala Gly Glu Gly Ser Gly Ser Pro
         -40   -35   -30
Val Arg Pro Pro Val Ser Thr Trp Gly Pro Ser Trp Ala Gln Leu Leu
         -25   -20   -15
Asp Ser Val Leu Trp Leu Gly Ala Leu Gly Leu Thr Ile Gln Ala Val
-10   -5   1   5
Phe Ser Thr Thr Gly Pro Ala Leu Leu Leu Leu Leu Val Ser Phe Leu
                10   15   20
Thr Phe Asp Leu Leu His Arg Pro Ala Val Thr Leu Cys His Ser Ala
         25   30   35
Asn Phe Ser Pro Gly Ala Arg Val Arg Gly Pro Val Lys Val Leu Asp
    40   45   50
Ser Arg Arg Leu Tyr Ser Cys Lys Trp Val Gln Ser Gln Asp Asn Leu
55   60   65   70
Ala Ser Arg Lys His Cys Cys Cys Cys Ser Trp Gly Trp Ala Arg Ser
                75   80   85

<210> 435
<211> 121

<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -16..-1

<400> 435
Met Glu Arg Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala
    -15   -10   -5
Ser Ala Gly Cys Ala Thr Thr Pro Ala Arg Asn Leu Ser Cys Tyr Gln
1   5   10   15
Cys Phe Lys Val Ser Ser Trp Thr Glu Cys Pro Pro Thr Trp Cys Ser
            20   25   30
Pro Leu Asp Gln Val Cys Ile Ser Asn Glu Val Val Val Ser Phe Ser
            35   40   45
Glu Ser Pro Pro Gly Arg Gly Xaa Val Pro Xaa Ala Gly Glu Xaa Pro
        50   55   60
Val Pro Pro Pro Leu Xaa Asp Leu Xaa Met Thr Pro Arg Xaa Xaa Arg
65   70   75   80
Ala Trp Gly Pro Val Gly Pro Lys Val Pro Pro Ala Val Ser Pro Ala
                85   90   95
Leu Gly Ser Gly Glu His Pro Xaa Xaa
            100   105

<210> 436
<211> 162
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -16..-1

<400> 436
Met Glu Arg Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala
    -15   -10   -5
Ser Ala Gly Cys Ala Thr Thr Pro Ala Arg Asn Leu Ser Cys Tyr Gln
1   5   10   15
Cys Phe Lys Val Ser Ser Trp Thr Glu Cys Pro Pro Thr Trp Cys Ser
            20   25   30
Pro Leu Asp Gln Val Cys Ile Ser Asn Glu Val Val Val Ser Phe Lys
            35   40   45
Trp Ser Val Arg Val Leu Leu Ser Lys Arg Cys Ala Pro Arg Cys Pro
        50   55   60
Asn Asp Asn Met Xaa Phe Glu Trp Ser Pro Ala Pro Met Val Gln Gly
65   70   75   80
Val Ile Thr Arg Arg Cys Cys Ser Trp Ala Leu Cys Asn Arg Ala Leu
                85   90   95
Thr Pro Gln Glu Gly Arg Trp Ala Leu Xaa Gly Gly Leu Leu Leu Gln
                100   105   110
Asp Pro Ser Arg Gly Xaa Lys Thr Trp Val Arg Pro Gln Leu Gly Leu
            115   120   125
Pro Leu Cys Leu Pro Xaa Ser Asn Pro Leu Cys Pro Xaa Glu Thr Gln
    130   135   140
Glu Gly
145

```
<210> 437
<211> 110
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -20..-1

<400> 437
Met Xaa Leu Met Val Leu Val Phe Thr Ile Gly Leu Thr Leu Leu Leu
-20                 -15                 -10                 -5
Gly Xaa Gln Ala Met Pro Ala Asn Arg Leu Ser Cys Tyr Arg Lys Ile
            1               5                   10
Leu Lys Asp His Asn Cys His Asn Leu Pro Glu Gly Val Ala Asp Leu
        15              20                  25
Thr Gln Ile Asp Val Asn Val Gln Asp His Phe Trp Asp Gly Lys Gly
        30              35                  40
Cys Glu Met Ile Cys Tyr Cys Asn Phe Lys Arg Ile Ala Leu Leu Pro
45              50                  55                  60
Lys Arg Arg Phe Leu Trp Thr Lys Asp Leu Phe Arg Asp Ser Leu Gln
                65              70                  75
Gln Ser Met Arg Ile Phe Met Tyr Ser Gly Glu His His Ser
            80              85                  90


<210> 438
<211> 71
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -15..-1

<400> 438
Met Lys Leu Leu Thr His Asn Leu Leu Ser Ser His Val Arg Gly Val
-15                 -10                 -5                  1
Gly Ser Arg Gly Phe Pro Leu Arg Leu Gln Ala Thr Glu Val Arg Ile
            5               10                  15
Cys Pro Val Glu Phe Asn Pro Asn Phe Val Ala Arg Met Ile Pro Lys
        20              25                  30
Val Glu Trp Ser Ala Phe Leu Glu Ala Xaa Asp Asn Leu Arg Leu Ile
    35              40                  45
Gln Val Pro Arg Arg Ala Gly
50              55


<210> 439
<211> 99
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -24..-1

<400> 439
Met Lys Ser Ala Lys Leu Gly Phe Leu Leu Arg Phe Phe Ile Phe Cys
            -20             -15                 -10
```

```
Ser Leu Asn Thr Leu Leu Leu Gly Gly Val Asn Lys Ile Ala Glu Lys
            -5                   1                   5
Ile Cys Gly Asp Leu Lys Asp Pro Cys Lys Leu Asp Met Asn Phe Gly
    10                  15                  20
Ser Cys Tyr Glu Val His Phe Arg Tyr Phe Tyr Asn Arg Thr Ser Lys
25                      30                  35                  40
Arg Cys Glu Thr Phe Val Phe Ser Ser Cys Asn Gly Asn Leu Asn Asn
                45                  50                  55
Phe Lys Leu Lys Ile Glu Arg Glu Val Xaa Cys Val Ala Lys Tyr Lys
            60                  65                  70
Pro Pro Arg
        75
```

```
<210> 440
<211> 169
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -25..-1

<400> 440
Met Arg Lys Pro Ala Ala Gly Phe Leu Pro Ser Leu Leu Lys Val Leu
-25                 -20                 -15                 -10
Leu Leu Pro Leu Ala Pro Ala Ala Ala Gln Asp Ser Thr Gln Ala Ser
            -5                   1                   5
Thr Pro Gly Ser Pro Leu Ser Pro Thr Glu Tyr Gln Arg Phe Phe Ala
        10                  15                  20
Leu Leu Thr Pro Thr Trp Lys Ala Glu Thr Thr Cys Arg Leu Arg Ala
    25                  30                  35
Thr His Gly Cys Arg Asn Pro Thr Leu Val Gln Leu Asp Gln Tyr Glu
40                  45                  50                  55
Asn His Gly Leu Val Pro Asp Gly Ala Val Cys Ser Asn Leu Pro Tyr
                60                  65                  70
Ala Ser Trp Phe Glu Ser Phe Cys Gln Phe Thr His Tyr Arg Cys Ser
            75                  80                  85
Asn His Val Tyr Tyr Ala Lys Arg Val Leu Cys Ser Gln Pro Val Ser
        90                  95                  100
Ile Leu Ser Pro Asn Thr Leu Lys Glu Ile Glu Xaa Ser Ala Glu Val
    105                 110                 115
Ser Pro Thr Thr Asp Asp Leu Pro His Leu Thr Pro Leu His Ser Asp
120                 125                 130                 135
Arg Thr Pro Asp Leu Pro Ala Leu Ala
                140
```

```
<210> 441
<211> 167
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -76..-1

<400> 441
Met Gly Asp Tyr Leu Leu Arg Gly Tyr Arg Met Leu Gly Glu Thr Cys
    -75                 -70                 -65
```

```
Ala Asp Cys Gly Thr Ile Leu Leu Gln Asp Lys Gln Arg Lys Ile Tyr
-60               -55               -50               -45
Cys Val Ala Cys Gln Glu Leu Asp Ser Asp Val Asp Lys Asp Asn Pro
              -40               -35               -30
Ala Leu Asn Ala Gln Ala Ala Leu Ser Gln Ala Arg Glu His Gln Leu
          -25               -20               -15
Ala Ser Ala Ser Glu Leu Pro Leu Gly Ser Arg Pro Ala Pro Gln Pro
          -10               -5                1
Pro Val Pro Arg Pro Glu His Cys Glu Gly Ala Ala Ala Gly Leu Lys
5                 10                15                20
Ala Ala Gln Gly Pro Pro Ala Pro Ala Val Pro Pro Asn Thr Xaa Val
              25                30                35
Met Ala Cys Thr Gln Thr Ala Leu Leu Gln Lys Leu Thr Trp Ala Ser
              40                45                50
Ala Glu Leu Gly Ser Xaa Thr Ser Xaa Gly Lys Xaa Ala Ser Ser Cys
          55                60                65
Val Ala Leu Ser Ala His Val Arg Arg Pro Cys Ala Ala Cys Ser Ser
          70                75                80
Tyr Ser Thr Lys Arg Ser Pro
85                90
```

```
<210> 442
<211> 70
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -15..-1

<400> 442
Met Ile Leu Cys Phe Leu Leu Pro His His Arg Leu Gln Glu Ala Arg
-15               -10               -5                1
Gln Ile Gln Val Leu Lys Met Leu Pro Arg Glu Lys Leu Arg Arg Arg
          5                 10                15
Glu Glu Arg Lys Gln Ile Asn Gly Lys Lys Xaa Arg Thr Lys Tyr Glu
          20                25                30
Thr Pro Arg Lys Xaa Xaa Gly Lys Lys Gly Gly Asn Xaa Xaa Xaa Xaa
          35                40                45
Xaa Leu Ser Lys Arg Asp
50                55
```

```
<210> 443
<211> 381
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -33..-1

<400> 443
Met Ser Trp Thr Val Pro Val Val Arg Ala Ser Gln Arg Val Ser Ser
          -30               -25               -20
Val Gly Ala Asn Xaa Leu Cys Leu Gly Met Ala Leu Cys Pro Arg Gln
          -15               -10               -5
Ala Thr Arg Ile Pro Leu Asn Gly Thr Trp Leu Phe Thr Pro Val Ser
          1                 5                 10                15
```

```
Lys Met Ala Thr Val Lys Ser Glu Leu Ile Glu Arg Phe Thr Ser Glu
             20                      25                      30
Lys Pro Val His His Ser Lys Val Ser Ile Ile Gly Thr Gly Ser Val
             35                      40                      45
Gly Met Ala Cys Ala Ile Ser Ile Leu Leu Lys Gly Leu Ser Asp Glu
         50                      55                      60
Leu Ala Leu Val Asp Leu Asp Glu Xaa Lys Leu Lys Gly Glu Thr Met
     65                      70                      75
Asp Leu Gln His Gly Ser Pro Phe Thr Lys Met Pro Asn Ile Val Cys
 80                      85                      90                      95
Ser Lys Xaa Tyr Phe Val Thr Ala Asn Ser Asn Leu Val Ile Ile Thr
                 100                     105                     110
Ala Gly Ala Arg Gln Xaa Lys Gly Glu Thr Arg Leu Asn Leu Xaa Gln
             115                     120                     125
Arg Asn Val Ala Ile Phe Lys Leu Met Ile Ser Ser Ile Val Gln Tyr
         130                     135                     140
Ser Pro His Cys Lys Leu Ile Ile Val Ser Asn Pro Val Asp Ile Leu
     145                     150                     155
Thr Tyr Val Ala Trp Lys Leu Ser Ala Phe Pro Lys Asn Arg Ile Ile
160                     165                     170                     175
Gly Ser Gly Cys Asn Leu Ile Xaa Ala Arg Phe Arg Phe Leu Ile Gly
                 180                     185                     190
Gln Lys Leu Gly Ile His Ser Glu Ser Cys His Gly Trp Ile Leu Gly
             195                     200                     205
Glu His Gly Asp Ser Ser Val Pro Val Trp Ser Gly Val Asn Ile Ala
         210                     215                     220
Gly Val Pro Leu Lys Asp Leu Asn Ser Asp Ile Gly Thr Asp Lys Asp
     225                     230                     235
Pro Glu Gln Trp Lys Asn Val His Lys Glu Val Thr Ala Thr Ala Tyr
240                     245                     250                     255
Glu Ile Ile Lys Met Lys Gly Tyr Thr Ser Trp Ala Ile Gly Leu Ser
                 260                     265                     270
Val Ala Asp Leu Thr Glu Ser Ile Leu Lys Asn Leu Arg Arg Ile His
             275                     280                     285
Pro Val Ser Thr Ile Thr Lys Gly Leu Tyr Gly Ile Xaa Glu Glu Val
             290                     295                     300
Phe Leu Ser Ile Pro Cys Ile Leu Gly Glu Asn Gly Ile Thr Asn Leu
     305                     310                     315
Ile Lys Ile Lys Leu Thr Pro Glu Glu Glu Ala His Leu Lys Lys Ser
320                     325                     330                     335
Ala Lys Thr Leu Trp Glu Ile Gln Asn Lys Leu Lys Leu
             340                     345
```

```
<210> 444
<211> 39
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -14..-1

<400> 444
Met Tyr Tyr Met Val Cys Leu Phe Phe Arg Leu Ile Phe Ser Glu His
                 -10                     -5                      1
Leu Pro Ile Ile Gly Thr Val Thr Ser His Lys Thr Gly Thr Leu Thr
         5                      10                      15
Val Tyr Pro Thr Ser Ala Gly
     20                      25
```

```
<210> 445
<211> 50
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -37..-1

<400> 445
Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val Asn
        -35                 -30                 -25
Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala Leu
    -20                 -15                 -10
Ser Pro Cys Leu Thr Ala Pro Lys Ser Pro Arg Leu Ala Met Met Pro
-5                   1                   5                   10
Asp Asn


<210> 446
<211> 51
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -26..-1

<400> 446
Met Thr Pro Trp Cys Leu Ala Cys Leu Gly Arg Arg Pro Leu Ala Ser
    -25                 -20                 -15
Leu Gln Trp Ser Leu Thr Leu Ala Trp Cys Gly Ser Gly Ser His Trp
-10                 -5                  1                   5
Thr Glu Arg Pro Xaa Gln Xaa Ser Pro Trp Xaa Ser Leu Ser Ala Thr
            10                  15                  20
Thr Arg Gly
        25


<210> 447
<211> 242
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -30..-1

<400> 447
Met Gly Glu Ala Ser Pro Pro Ala Pro Ala Arg Arg His Leu Leu Val
-30                 -25                 -20                 -15
Leu Leu Leu Leu Leu Ser Thr Leu Val Ile Pro Ser Ala Ala Ala Pro
                -10                 -5                  1
Ile His Asp Ala Asp Ala Gln Glu Ser Ser Leu Gly Leu Thr Gly Leu
        5                   10                  15
Gln Ser Leu Leu Gln Gly Phe Ser Arg Leu Phe Leu Lys Gly Asn Leu
    20                  25                  30
Leu Arg Gly Ile Asp Ser Leu Phe Ser Ala Pro Met Asp Phe Arg Gly
```

```
35                        40                        45                        50
Leu Pro Gly Asn Tyr His Lys Glu Glu Asn Gln Glu His Gln Leu Gly
              55                        60                        65
Asn Asn Thr Leu Ser Ser His Leu Gln Ile Asp Lys Met Thr Asp Asn
              70                        75                        80
Lys Thr Gly Glu Val Leu Ile Ser Glu Asn Val Val Ala Ser Ile Gln
              85                        90                        95
Pro Xaa Glu Gly Xaa Phe Glu Gly Asp Leu Lys Val Pro Arg Met Glu
      100                        105                       110
Glu Lys Glu Ala Leu Val Pro Xaa Gln Lys Ala Thr Asp Ser Phe His
115                        120                       125                       130
Thr Glu Leu His Pro Arg Val Ala Phe Trp Ile Ile Lys Leu Pro Arg
              135                       140                       145
Arg Arg Ser His Gln Asp Ala Leu Glu Gly Gly His Trp Leu Xaa Glu
              150                       155                       160
Lys Arg His Arg Leu Gln Ala Ile Arg Asp Gly Leu Arg Lys Gly Thr
              165                       170                       175
His Lys Asp Xaa Leu Xaa Xaa Gly Thr Glu Ser Ser Ser His Ser Arg
      180                       185                       190
Leu Ser Pro Arg Lys Xaa His Leu Leu Tyr Ile Leu Xaa Pro Ser Arg
195                       200                       205                       210
Gln Leu
```

```
<210> 448
<211> 154
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -60..-1

<400> 448
Met Gly Ser Lys Cys Cys Lys Gly Gly Pro Asp Glu Asp Ala Val Glu
-60                  -55                  -50                       -45
Arg Gln Arg Arg Gln Lys Leu Leu Leu Ala Gln Leu His His Arg Lys
              -40                        -35                       -30
Arg Val Lys Ala Ala Gly Gln Ile Gln Ala Trp Trp Arg Gly Val Leu
              -25                        -20                       -15
Val Arg Arg Thr Leu Leu Val Ala Ala Leu Arg Ala Trp Met Ile Gln
              -10                        -5                         1
Cys Trp Trp Arg Thr Leu Val Gln Arg Arg Ile Arg Gln Arg Arg Gln
5                    10                        15                        20
Ala Leu Leu Gly Val Tyr Val Ile Gln Glu Gln Ala Ala Val Lys Leu
              25                        30                        35
Gln Ser Cys Ile Arg Met Trp Gln Cys Arg Gln Cys Tyr Arg Gln Met
              40                        45                        50
Cys Asn Ala Leu Cys Leu Phe Gln Val Pro Lys Ser Ser Leu Ala Phe
              55                        60                        65
Gln Thr Asp Gly Phe Leu Gln Val Gln Tyr Ala Ile Pro Ser Lys Gln
              70                        75                        80
Pro Glu Phe His Ile Glu Ile Leu Ser Ile
85                        90
```

```
<210> 449
<211> 89
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -61..-1

<400> 449
Met Asn Ala Ala Ile Asn Thr Gly Pro Ala Pro Ala Val Thr Lys Thr
    -60                 -55             -50
Glu Thr Glu Val Gln Asn Pro Asp Val Leu Trp Asp Leu Asp Ile Pro
-45                 -40             -35                     -30
Glu Ala Arg Ser His Ala Asp Gln Asp Ser Asn Pro Lys Ala Glu Ala
                -25             -20                     -15
Leu Leu Pro Cys Asn Leu His Cys Ser Trp Leu His Ser Ser Pro Arg
            -10             -5                  1
Pro Asp Pro His Ser His Phe Pro Ser Xaa Arg Arg Cys Pro Leu Pro
    5               10                  15
His Pro Cys Ala Thr Tyr Pro Pro Xaa
20              25
```

```
<210> 450
<211> 73
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -26..-1

<400> 450
Met Arg Met Ser Leu Ala Gln Arg Val Leu Leu Thr Trp Leu Phe Thr
    -25                 -20             -15
Leu Leu Phe Leu Ile Met Leu Val Leu Lys Leu Asp Glu Lys Ala Pro
-10             -5                  1                   5
Trp Asn Trp Phe Leu Ile Phe Ile Pro Val Trp Ile Phe Asp Thr Ile
                10              15                  20
Leu Leu Val Leu Leu Ile Val Lys Met Ala Gly Arg Cys Lys Ser Gly
        25              30                  35
Phe Asp Leu Asp Met Asp His Thr Ile
    40              45
```

```
<210> 451
<211> 54
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -34..-1

<400> 451
Met Ile Pro Leu Ile Ser His Leu Ala Glu Ala Ala Pro Pro Thr Ser
                -30             -25             .       -20
Trp Ser Leu Ile Ser Ser Val Leu Asn Val Gly His Leu Leu Phe Ser
            -15             -10                 -5
Ser Ala Cys Ser Val Ser Leu Glu Ala Leu Ser Thr Arg Asn Ile Lys
            1               5                   10
Ala Ile Ile Leu Met Lys
15              20
```

<210> 452
<211> 121
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -38..-1

<400> 452
Met Glu Ser Pro Gln Leu His Cys Ile Leu Asn Ser Asn Ser Val Ala
         -35   -30   -25
Cys Ser Phe Ala Val Gly Ala Gly Phe Leu Ala Phe Leu Ser Cys Leu
         -20   -15   -10
Ala Phe Leu Val Leu Asp Thr Gln Glu Thr Arg Ile Ala Gly Thr Arg
         -5   1   5   10
Phe Lys Thr Ala Phe Gln Leu Leu Asp Phe Ile Leu Ala Val Leu Trp
              15   20   25
Ala Val Val Trp Phe Met Gly Phe Cys Phe Leu Ala Asn Gln Trp Gln
              30   35   40
His Ser Pro Pro Lys Glu Xaa Leu Leu Gly Ser Ser Ser Ala Gln Ala
         45   50   55
Ala Ile Gly Xaa His Leu Leu Leu His Pro Cys Leu Asp Ile Pro Xaa
         60   65   70
Leu Pro Gly Xaa Pro Gly Pro Pro Lys
75   80

<210> 453
<211> 166
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -37..-1

<400> 453
Met Ser Thr Val Gly Leu Phe His Phe Pro Thr Pro Leu Thr Arg Ile
         -35   -30   -25
Cys Pro Ala Pro Trp Gly Leu Arg Leu Trp Glu Lys Leu Thr Leu Leu
         -20   -15   -10
Ser Pro Gly Ile Ala Val Thr Pro Val Gln Met Ala Gly Lys Lys Asp
-5   1   5   10
Tyr Pro Ala Leu Leu Ser Leu Asp Glu Asn Glu Leu Glu Glu Gln Phe
              15   20   25
Val Lys Gly His Gly Pro Gly Gly Gln Ala Thr Asn Lys Thr Ser Asn
              30   35   40
Cys Val Val Leu Lys Xaa Ile Pro Ser Gly Ile Val Val Lys Cys His
         45   50   55
Gln Thr Arg Ser Val Asp Gln Asn Arg Lys Leu Ala Arg Lys Ile Leu
60   65   70   75
Gln Glu Lys Val Xaa Val Phe Tyr Asn Gly Glu Asn Ser Pro Val His
              80   85   90
Lys Glu Lys Arg Glu Ala Ala Lys Lys Lys Gln Glu Arg Lys Lys Arg
              95   100   105
Ala Lys Glu Thr Leu Glu Lys Lys Xaa Leu Leu Lys Xaa Leu Trp Glu
         110   115   120

```
Ser Ser Lys Lys Val His
    125
```

```
<210> 454
<211> 180
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -26..-1

<400> 454
Met Gly Ile Gln Thr Ser Pro Val Leu Leu Ala Ser Leu Gly Val Gly
    -25                   -20                   -15
Leu Val Thr Leu Leu Gly Leu Ala Val Gly Ser Tyr Leu Val Arg Arg
-10                   -5                    1                     5
Ser Arg Arg Pro Gln Val Thr Leu Leu Asp Pro Asn Glu Lys Tyr Leu
             10                   15                    20
Leu Arg Leu Leu Asp Lys Thr Thr Val Ser His Asn Thr Lys Arg Phe
    25                   30                    35
Arg Phe Ala Leu Pro Thr Ala His His Thr Leu Gly Leu Pro Val Gly
    40                   45                    50
Lys His Ile Tyr Leu Ser Thr Arg Ile Asp Gly Ser Leu Val Ile Arg
55                   60                    65                    70
Pro Tyr Thr Pro Val Thr Ser Asp Glu Asp Gln Gly Tyr Val Asp Leu
             75                   80                    85
Val Xaa Lys Val Tyr Leu Lys Gly Val His Pro Lys Phe Pro Glu Gly
             90                   95                    100
Gly Lys Met Ser Xaa Tyr Leu Asp Xaa Leu Lys Val Gly Asp Xaa Val
    105                   110                   115
Glu Phe Xaa Gly Pro Ser Gly Leu Leu Thr Tyr Thr Gly Lys Gly His
    120                   125                   130
Phe Asn Ile Gln Pro Asn Lys Asn Leu His Gln Asn Pro Glu Trp Arg
135                   140                   145                   150
Arg Asn Trp Glu
```

```
<210> 455
<211> 91
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -64..-1

<400> 455
Met Thr Pro Arg Ile Leu Ser Glu Val Gln Phe Ser Ala Phe Cys Pro
             -60                   -55                   -50
Tyr Trp Thr Ile Ala Arg Ile Leu Glu Arg Val Gly Ser Ala Cys Phe
             -45                   -40                   -35
Arg Leu Glu Leu Cys Ala Ala Ile Val Gly Tyr Phe Val Leu Asp Val
             -30                   -25                   -20
Arg Thr Phe Leu Phe Ile Val Val Cys Val Ile Cys Val Thr Leu Asn
             -15                   -10                   -5
Phe Pro Arg Phe Tyr Phe Leu Cys Leu Ser Ser Leu Thr Ala Phe Gly
1                     5                     10                    15
Thr Pro Pro Ile Gly Val His Ile Pro Ser Pro
```

20 25

```
<210> 456
<211> 257
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -23..-1

<400> 456
Met Arg Arg Ile Ser Leu Thr Ser Ser Pro Val Arg Leu Leu Leu Xaa
              -20                -15                -10
Leu Leu Leu Leu Leu Ile Ala Leu Glu Ile Met Val Gly Gly His Ser
          -5                  1                5
Leu Cys Phe Asn Phe Thr Ile Lys Ser Leu Ser Arg Pro Gly Gln Pro
10               15               20               25
Trp Cys Glu Ala His Val Phe Leu Asn Lys Asn Leu Phe Leu Gln Tyr
              30               35               40
Asn Ser Asp Asn Asn Met Val Lys Pro Leu Gly Leu Leu Gly Lys Lys
          45               50               55
Val Tyr Ala Thr Ser Thr Trp Gly Glu Leu Thr Gln Thr Leu Gly Glu
          60               65               70
Val Gly Arg Asp Leu Arg Met Leu Leu Cys Asp Ile Lys Pro Gln Ile
      75               80               85
Lys Thr Ser Asp Pro Ser Thr Leu Gln Val Xaa Xaa Phe Cys Gln Arg
90               95               100              105
Glu Ala Glu Arg Cys Thr Gly Ala Ser Trp Gln Phe Ala Thr Asn Gly
              110              115              120
Glu Lys Ser Leu Leu Phe Asp Ala Met Asn Met Thr Trp Thr Val Ile
          125              130              135
Asn His Glu Ala Ser Xaa Ile Lys Glu Thr Trp Lys Lys Asp Arg Xaa
          140              145              150
Leu Glu Xaa Tyr Phe Arg Lys Leu Ser Lys Gly Asp Cys Asp His Trp
      155              160              165
Leu Arg Glu Phe Leu Gly His Trp Glu Ala Met Pro Xaa Pro Xaa Val
170              175              180              185
Ser Pro Xaa Asn Ala Ser Xaa Ile His Trp Ser Ser Ser Xaa Leu Pro
              190              195              200
Xaa Xaa Trp Ile Ile Leu Gly Ala Phe Ile Leu Leu Xaa Leu Met Gly
          205              210              215
Ile Val Leu Ile Cys Val Trp Trp Gln Asn Gly Xaa Xaa Ser Thr Xaa
          220              225              230
Xaa
```

```
<210> 457
<211> 193
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -60..-1

<400> 457
Met Cys Pro Ser Leu Glu Glu Ala Pro Ser Val Lys Gly Thr Leu Pro
-60               -55              -50              -45
```

```
Cys Ser Gly Gln Gln Gln Pro Phe Pro Phe Gly Ala Ser Asn Ile Pro
              -40              -35                  -30
Leu Leu Leu Gly Arg Ser Arg Lys Val Ala Arg Gly Ala Pro Val Leu
              -25              -20                  -15
Trp Pro Phe Leu Thr Trp Ile Asn Pro Ala Leu Ser Ile Cys Asp Pro
              -10               -5                  1
Leu Gly Ser Cys Gly Trp Xaa Cys His Thr Ala Gln Val Pro Ala Pro
 5               10               15                  20
Leu Gln Leu Pro Thr Ala Cys Pro Pro Leu Pro His Gly Thr Arg Ala
              25               30                  35
Val Gly Pro Thr Pro Gly Leu Leu Pro Glu Ala Ala Ala Pro Xaa Thr
              40               45                  50
Xaa Gly Ala Leu Ser Ser Arg Ser Arg His Trp Ser Cys Ser Ile Val
 55               60                  65
Xaa Cys Leu His Leu His Xaa Leu Leu Ser Val Glu Thr Arg Xaa Phe
 70               75                  80
Xaa Lys His Leu Leu Val Leu Leu Val Ala Val Ala His Ser Val Leu
 85                    90               95                  100
Glu Pro Pro Ala Leu Val Pro Asn Val Gln Cys Glu Met Cys Thr His
              105              110                  115
Ser Gly Pro Arg Asp Leu Glu Ala Ala Val Val Ser Pro Ala Pro Trp
              120              125                  130
Glu
```

```
<210> 458
<211> 107
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -28..-1

<400> 458
Met Val Leu Thr Leu Gly Glu Ser Trp Pro Val Leu Val Gly Arg Arg
              -25              -20                  -15
Phe Leu Ser Leu Ser Ala Ala Asp Gly Ser Asp Gly Ser His Asp Ser
              -10               -5                  1
Trp Asp Val Glu Arg Val Ala Glu Trp Pro Trp Leu Ser Gly Thr Ile
 5               10               15                  20
Arg Ala Val Ser His Thr Asp Val Thr Lys Lys Asp Leu Lys Val Cys
              25               30                  35
Val Glu Phe Xaa Gly Glu Ser Trp Arg Lys Arg Arg Trp Ile Glu Val
              40               45                  50
Tyr Ser Leu Leu Arg Lys Ala Phe Leu Val Lys His Asn Leu Val Leu
              55               60                  65
Ala Glu Arg Lys Ser Pro Glu Ile Ser Trp Gly
 70               75
```

```
<210> 459
<211> 121
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -13..-1
```

```
<400> 459
Met Leu Val Leu Arg Ser Ala Leu Thr Arg Ala Leu Ala Ser Arg Thr
              -10                   -5                   1
Leu Ala Pro Gln Met Cys Ser Ser Phe Ala Thr Gly Pro Arg Gln Tyr
     5               10                   15
Asp Gly Ile Phe Tyr Glu Phe Arg Ser Tyr Tyr Leu Lys Pro Ser Lys
20               25                   30                   35
Met Asn Glu Phe Leu Glu Asn Phe Glu Lys Asn Ala Gln Leu Arg Thr
              40                   45                   50
Ala His Ser Glu Leu Val Gly Tyr Trp Ser Val Xaa Phe Gly Gly Arg
         55                   60                   65
Met Xaa Thr Val Phe His Ile Trp Lys Tyr Asp Asn Phe Ala His Arg
         70                   75                   80
Thr Glu Phe Gln Lys Ala Leu Ala Lys Asp Lys Glu Trp Gln Glu Gln
     85                   90                   95
Phe Leu Ile Pro Asn Leu Ala Leu Asn
100                   105


<210> 460
<211> 44
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -17..-1

<400> 460
Met Lys Val Gly Val Leu Trp Leu Ile Ser Phe Phe Thr Phe Thr Asp
              -15                   -10                   -5
Gly His Gly Gly Phe Leu Gly Val Ser Trp Cys Tyr Val Ser Tyr Leu
     1               5                   10                   15
Phe Ser Thr Asn Ser Pro Leu Ser Phe Arg Arg Ile
              20                   25


<210> 461
<211> 109
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -13..-1

<400> 461
Met Cys Leu Leu Thr Ala Leu Val Thr Gln Val Ile Ser Leu Arg Lys
              -10                   -5                   1
Asn Ala Glu Arg Thr Cys Leu Cys Lys Arg Arg Trp Pro Trp Xaa Pro
     5               10                   15
Ser Pro Arg Ile Tyr Cys Ser Ser Thr Pro Cys Asp Ser Lys Phe Pro
20               25                   30                   35
Thr Val Tyr Ser Ser Ala Pro Phe His Ala Pro Leu Pro Val Gln Asn
              40                   45                   50
Ser Leu Trp Gly His Pro Leu His Gly Cys Ser Trp Gln Cys His His
         55                   60                   65
Pro Gln Gly Gln Asn Leu Gln Pro Ala Ser Leu Xaa Thr His Leu Ser
         70                   75                   80
Lys Pro Lys Arg His Phe Xaa Lys Lys Xaa Cys Gln Ala
```

        85   90   95

<210> 462
<211> 143
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -41..-1

<400> 462
Met Ala Thr Ala Thr Glu Gln Trp Val Leu Val Glu Met Val Gln Ala
     -40   -35   -30
Leu Tyr Glu Ala Pro Ala Tyr His Leu Ile Leu Glu Gly Ile Leu Ile
-25   -20   -15   -10
Leu Trp Ile Ile Arg Leu Leu Phe Ser Lys Thr Tyr Lys Leu Gln Glu
          -5   1   5
Arg Ser Asp Leu Thr Val Lys Glu Lys Glu Glu Leu Ile Glu Glu Trp
     10   15   20
Gln Pro Glu Pro Leu Val Pro Pro Val Pro Lys Asp His Pro Ala Leu
     25   30   35
Asn Tyr Asn Ile Val Ser Gly Pro Pro Ser His Lys Thr Val Val Asn
40   45   50   55
Gly Lys Glu Cys Ile Asn Phe Ala Ser Phe Asn Phe Leu Gly Leu Leu
          60   65   70
Asp Asn Pro Arg Val Lys Ala Ala Ala Leu Ala Ser Leu Lys Lys Tyr
          75   80   85
Gly Val Gly Thr Cys Gly Pro Cys Gly Phe Tyr Gly Thr Phe Glu
          90   95   100

<210> 463
<211> 232
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -30..-1

<400> 463
Met Ala Ala Thr Ser Gly Thr Asp Glu Pro Val Ser Gly Glu Leu Val
-30   -25   -20   -15
Ser Val Ala His Ala Leu Ser Leu Pro Ala Glu Ser Tyr Gly Asn Xaa
         -10   -5   1
Xaa Asp Ile Glu Met Ala Trp Ala Met Arg Ala Met Gln His Ala Glu
      5   10   15
Val Tyr Tyr Lys Leu Ile Ser Ser Val Asp Pro Gln Phe Leu Lys Leu
     20   25   30
Thr Lys Val Asp Asp Gln Ile Tyr Ser Glu Phe Arg Lys Asn Phe Glu
35   40   45   50
Thr Leu Arg Ile Asp Val Leu Xaa Pro Glu Xaa Leu Lys Ser Glu Ser
          55   60   65
Ala Lys Glu Pro Pro Gly Tyr Asn Ser Leu Pro Leu Lys Leu Leu Gly
          70   75   80
Thr Gly Lys Ala Ile Thr Lys Leu Phe Ile Ser Val Phe Arg Thr Lys
          85   90   95
Lys Glu Arg Lys Glu Ser Thr Met Glu Glu Lys Lys Glu Leu Thr Val

```
         100                      105                      110
Glu Lys Lys Arg Thr Pro Arg Met Glu Glu Arg Lys Glu Leu Ile Val
115                      120                      125                      130
Glu Lys Lys Lys Arg Lys Glu Ser Thr Glu Lys Thr Lys Leu Thr Lys
                 135                      140                      145
Glu Glu Lys Lys Gly Lys Lys Leu Thr Lys Lys Ser Thr Lys Val Val
                 150                      155                      160
Lys Lys Leu Cys Lys Val Tyr Arg Glu Gln His Ser Arg Ser Tyr Asp
             165                      170                      175
Ser Ile Glu Thr Thr Ser Thr Thr Val Leu Leu Ala Gln Thr Pro Leu
         180                      185                      190
Val Lys Cys Lys Phe Leu Tyr Asn
195                      200
```


<210> 464
<211> 61
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 464
```
Met Thr Phe Arg His Gln Asp Asn Ser Leu Met Phe Phe Ser Met Met
     -20                      -15                      -10
Ala Thr Cys Thr Ser Asn Val Gly Phe Thr His Thr Thr Met Asn Cys
-5                        1                  5                        10
Ser Leu Thr Ser Pro Val Asp Phe Lys Asp Leu Leu Arg Val Leu Leu
             15                      20                      25
Ile Lys Phe Gly Tyr Asp Arg Lys Ser Thr Ile Lys Ser
         30                      35                      40
```


<210> 465
<211> 34
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -19..-1

<400> 465
```
Met Phe Leu Lys Ser Gly Ala Gly Leu Ser Ser Cys Leu Leu Pro Leu
                 -15                      -10                      -5
Cys Trp Leu Glu Arg Lys Asp His Gly Arg Arg Pro Ser Xaa His Pro
             1                  5                        10
Gly Arg
     15
```


<210> 466
<211> 215
<212> PRT
<213> Homo sapiens

<220>

<221> SIGNAL
<222> -54..-1

<400> 466
Met Asn Xaa Tyr Ala Ser Pro Phe Asn Xaa Gln Leu Xaa Tyr Leu Xaa
            -50   -45   -40
Leu Ser Arg Phe Glu Cys Val His Arg Asp Gly Arg Val Ile Thr Leu
            -35   -30   -25
Ser Tyr Gln Glu Gln Glu Leu Gln Asp Phe Leu Leu Ser Gln Met Ser
            -20   -15   -10
Gln His Gln Val His Ala Val Gln Gln Leu Ala Lys Val Met Gly Trp
        -5   1   5   10
Gln Val Leu Ser Phe Ser Asn His Val Gly Leu Gly Pro Ile Glu Ser
                15   20   25
Xaa Gly Asn Ala Ser Ala Ile Thr Val Ala Pro Gln Val Val Thr Met
            30   35   40
Leu Phe Gln Phe Val Met Asp Leu Lys Val Ala Ala Arg Leu Trp Phe
            45   50   55
Ser Phe Leu Val Thr Asn Val Lys Thr Phe Gln Lys Val Met Phe Tyr
        60   65   70
Lys Ile Thr Asn Gly Val Ile Phe Val Gly His Ser Lys Lys Phe Ser
75   80   85   90
Gly Ile Lys Trp Lys Val Xaa Ile Leu Phe Ile Lys Trp Xaa Cys Leu
                95   100   105
Cys Leu His Leu Ala Leu Val Tyr Tyr Asp Phe Phe Gln Met Phe Pro
            110   115   120
Lys Xaa Val Ser Xaa Asn Phe Asp Leu Lys Cys Leu Gln Ile Asn Tyr
            125   130   135
Lys His Lys Glu Glu Ile Thr Ser Lys Arg Val Leu Phe Leu Lys Ile
            140   145   150
Ile Ile Arg Lys Cys Phe Ile
155   160

<210> 467
<211> 27
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -17..-1

<400> 467
Met Val Val His Leu Leu Tyr Ala His Leu Ser Phe Thr Ser Lys Arg
            -15   -10   -5
Ala Val Val Met Leu Lys Leu Glu Ile Thr Phe
    1   5   10

<210> 468
<211> 85
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -24..-1

<400> 468

```
Met Cys Ser His Ala Ser Met Ser Phe His Thr Leu Phe His Leu Leu
            -20                      -15                      -10
Phe Leu Pro His Tyr Ile Glu Thr Phe Lys Pro Gln Ser Lys His Cys
             -5                       1                   5
Phe Phe Trp Ile Ala Ala Phe Leu Thr Ser Leu Leu Thr Pro Gln Ser
    10                      15                      20
Leu Gln Gly Phe His Ser Ser Leu Cys Ala Leu Arg Ser Gln His Phe
25                      30                      35                      40
Pro Ser Thr Cys Asn Cys Phe Cys Tyr Leu Thr Ile Ile Ala Leu Xaa
                45                      50                      55
Tyr Trp Asp Asn Leu
            60
```

```
<210> 469
<211> 51
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -16..-1

<400> 469
Met Leu Arg Ile Ala Leu Thr Leu Ile Pro Ser Met Leu Ser Arg Ala
    -15                     -10                     -5
Ala Gly Trp Cys Trp Tyr Lys Glu Pro Thr Gln Gln Phe Ser Tyr Leu
1               5                       10                      15
Cys Leu Pro Cys Leu Ser Trp Asn Lys Lys Gly Asn Val Leu Gln Leu
                20                      25                      30
Pro Asn Phe
        35
```

```
<210> 470
<211> 67
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -43..-1

<400> 470
Met Thr Pro Gln Tyr Leu Pro His Gly Gly Lys Tyr Gln Val Leu Gly
            -40                     -35                     -30
Asp Tyr Ser Leu Ala Val Val Phe Pro Leu His Phe Ser Asp Leu Ile
            -25                     -20                     -15
Ser Val Leu Tyr Leu Ile Pro Lys Thr Leu Thr Thr Asn Thr Ala Val
            -10                     -5                      1               5
Lys His Ser Ile Gln Lys Asn Cys Met Xaa Leu Val Leu Gly Lys Leu
                    10                      15                      20
Leu Ser Gln
```

```
<210> 471
<211> 63
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -15..-1

<400> 471
Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe Ala Arg Ala Leu
-15                 -10                 -5                   1
Asp Gly Cys Arg Asn Gly Ile Ala His Pro Ala Ser Glu Lys His Arg
          5                   10                  15
Leu Glu Lys Cys Arg Glu Leu Glu Ser Ser His Ser Ala Pro Gly Ser
          20                  25                  30
Thr Gln His Arg Arg Lys Thr Thr Arg Arg Asn Tyr Ser Ser Ala
          35                  40                  45
```

```
<210> 472
<211> 179
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -58..-1

<400> 472
Met Ser Thr Gly Gln Leu Tyr Arg Met Glu Asp Ile Gly Arg Phe His
              -55           -50               -45
Ser Gln Gln Pro Gly Ser Leu Thr Pro Ser Ser Pro Thr Val Gly Glu
          -40               -35               -30
Ile Ile Tyr Asn Asn Thr Arg Asn Thr Leu Gly Trp Ile Gly Gly Ile
          -25               -20               -15
Leu Met Gly Ser Phe Gln Gly Thr Ile Ala Gly Gln Gly Thr Gly Ala
-10               -5                   1                   5
Thr Ser Ile Ser Glu Leu Cys Lys Gly Gln Glu Leu Glu Pro Ser Gly
          10                  15                  20
Ala Gly Leu Thr Val Ala Pro Pro Gln Ala Val Ser Leu Gln Gly Ile
          25                  30                  35
Tyr Thr Leu Pro Trp Leu Leu Gln Leu Phe His Ser Thr Ala Leu Xaa
          40                  45                  50
Xaa Xaa Gln Gln Pro Asn Gly Ser Leu Ser Leu Asn Ile Ser Ser Ser
55                  60                  65                  70
His Ala Pro Xaa Pro Xaa Thr Cys Thr Leu Glu Pro Gly Val Asp Pro
              75                  80                  85
Thr Arg Xaa Val Cys Ile Asn Pro His Pro Pro Pro Ile Leu Lys
          90                  95                  100
Xaa Pro Leu Ser Pro Tyr Pro Lys Pro Gln Leu Gly Thr His Ala Gly
          105                 110                 115
Gln Val Asn
    120
```

```
<210> 473
<211> 238
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -71..-1
```

<400> 473
Met Xaa Xaa Phe Thr Asp Pro Ser Ser Val Asn Glu Lys Lys Arg Arg
    -70   -65   -60
Glu Arg Glu Glu Arg Gln Asn Ile Val Leu Trp Arg Gln Pro Leu Ile
-55   -50   -45   -40
Thr Leu Gln Tyr Phe Ser Leu Glu Ile Leu Val Ile Leu Lys Glu Trp
            -35   -30   -25
Thr Ser Lys Leu Trp His Arg Gln Ser Ile Val Val Ser Phe Leu Leu
            -20   -15   -10
Leu Leu Ala Gly Leu Ile Ala Thr Tyr Tyr Val Glu Gly Val His Gln
         -5   1   5
Gln Tyr Val Gln Arg Ile Glu Lys Gln Phe Leu Leu Tyr Ala Tyr Trp
10   15   20   25
Ile Gly Leu Gly Ile Leu Ser Ser Val Gly Leu Gly Thr Gly Leu His
            30   35   40
Thr Phe Leu Leu Tyr Leu Gly Pro His Ile Ala Ser Val Thr Leu Ala
            45   50   55
Ala Tyr Glu Cys Asn Ser Val Asn Phe Pro Glu Pro Pro Tyr Pro Asp
         60   65   70
Gln Ile Ile Cys Pro Asp Glu Glu Gly Thr Glu Gly Thr Ile Ser Leu
75   80   85
Trp Ser Ile Ile Ser Lys Val Arg Ile Glu Ala Cys Met Trp Gly Ile
90   95   100   105
Gly Thr Ala Ile Gly Glu Leu Pro Pro Tyr Phe Met Ala Arg Ala Ala
            110   115   120
Arg Leu Ser Gly Ala Glu Pro Asp Asp Glu Glu Tyr Gln Glu Phe Glu
            125   130   135
Glu Met Leu Glu His Ala Glu Ser Ala Gln Val Arg Thr Val Gly Ile
            140   145   150
Glu Asn Arg Thr Leu Tyr Phe Phe Leu Lys Arg Leu Leu Arg
         155   160   165

<210> 474
<211> 178
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -37..-1

<400> 474
Met Glu Arg Gln Ser Arg Val Met Ser Glu Lys Asp Glu Tyr Gln Phe
            -35   -30   -25
Gln His Gln Gly Ala Val Glu Leu Leu Val Phe Asn Phe Leu Leu Ile
         -20   -15   -10
Leu Thr Ile Leu Thr Ile Trp Leu Phe Lys Asn His Arg Phe Arg Phe
-5   1   5   10
Leu His Glu Thr Gly Gly Ala Met Val Tyr Gly Leu Xaa Met Gly Leu
         15   20   25
Ile Leu Xaa Tyr Ala Thr Ala Pro Thr Asp Ile Glu Ser Gly Xaa Val
         30   35   40
Tyr Asp Cys Val Lys Leu Thr Phe Ser Pro Ser Thr Leu Leu Val Asn
      45   50   55
Ile Thr Asp Gln Val Tyr Glu Tyr Lys Tyr Lys Arg Glu Ile Ser Gln
60   65   70   75
His Xaa Ile Asn Pro His Xaa Gly Asn Ala Ile Leu Glu Lys Met Thr
            80   85   90
Phe Asp Pro Xaa Ile Phe Phe Asn Val Leu Leu Pro Pro Ile Ile Phe

```
                     95                      100                     105
His Ala Gly Tyr Ser Leu Lys Lys Arg His Phe Phe Gln Asn Leu Gly
        110                      115                     120
Ser Ile Leu Thr Tyr Ala Phe Leu Gly Thr Ala Ile Ser Cys Ile Val
        125                      130                     135
Ile Gly
140
```

```
<210> 475
<211> 96
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 475
Met Ser Met Gln Phe Leu Phe Lys Met Val Ala Leu Cys Cys Cys Leu
        -20                     -15                     -10
Trp Lys Ile Ser Gly Cys Glu Glu Val Pro Leu Thr Tyr Asn Leu Leu
-5                       1                       5                       10
Lys Cys Leu Leu Asp Lys Ala His Cys Val Leu Leu Thr Pro Cys Gly
            15                      20                      25
Tyr Ile Phe Ser Leu Ile Ser Pro Glu Ile Leu Lys Leu Thr Leu Ile
        30                      35                      40
Thr Leu Xaa Ile Leu Leu Ile Leu Lys Asn Leu His Leu Leu Trp Leu
        45                      50                      55
Thr Val Ser Ser Xaa Cys Val His Arg Ser Ser Ala Arg Lys Glu Lys
60                      65                      70                      75
```

```
<210> 476
<211> 41
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -24..-1

<400> 476
Met His Thr Phe Ala Asn Asp Arg Gly Leu Tyr Arg Ile Leu Leu Leu
                -20                     -15                     -10
His Phe Tyr Cys Leu Leu Arg Ser Ser Glu Tyr Ile Leu Gly Tyr Lys
            -5                      1                       5
Val Leu Gly Val Phe Phe Pro Ile Leu
        10                      15
```

```
<210> 477
<211> 113
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -27..-1
```

<400> 477
Met Arg Xaa Lys Trp Lys Met Gly Gly Met Lys Tyr Ile Phe Ser Leu
        -25   -20   -15
Leu Phe Phe Leu Leu Leu Glu Gly Gly Xaa Thr Glu Gln Val Xaa His
        -10   -5   1   5
Ser Glu Thr Tyr Cys Met Phe Gln Asp Lys Lys Tyr Arg Val Gly Glu
                 10   15   20
Arg Trp His Pro Tyr Leu Glu Pro Tyr Gly Leu Val Tyr Cys Val Asn
             25   30   35
Cys Ile Cys Ser Glu Asn Gly Asn Val Leu Cys Ser Arg Val Arg Cys
         40   45   50
Pro Asn Val His Cys Leu Ser Pro Val His Ile Pro His Leu Cys Cys
     55   60   65
Pro Arg Cys Pro Glu Asp Ser Leu Pro Pro Val Asn Asn Xaa Val Thr
70   75   80   85
Ser

<210> 478
<211> 250
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -18..-1

<400> 478
Met Arg Ile Leu Gln Leu Ile Leu Leu Ala Leu Ala Thr Gly Leu Val
        -15   -10   -5
Gly Gly Glu Thr Arg Ile Ile Lys Gly Phe Glu Cys Lys Pro His Ser
         1   5   10
Gln Pro Trp Gln Ala Ala Leu Phe Glu Lys Thr Arg Leu Leu Cys Gly
15   20   25   30
Ala Thr Leu Ile Ala Pro Arg Trp Leu Leu Thr Ala Ala His Cys Leu
             35   40   45
Lys Pro Arg Tyr Ile Xaa His Leu Gly Gln His Asn Leu Gln Lys Glu
             50   55   60
Glu Gly Cys Glu Gln Thr Arg Thr Ala Thr Glu Ser Phe Pro His Pro
         65   70   75
Gly Phe Asn Asn Ser Leu Pro Asn Lys Asp Xaa Xaa Asn Asp Ile Met
     80   85   90
Leu Val Xaa Met Xaa Ser Pro Val Ser Ile Thr Trp Ala Val Arg Pro
95   100   105   110
Leu Thr Leu Ser Ser Arg Cys Val Thr Ala Gly Thr Ser Cys Leu Ile
             115   120   125
Ser Gly Trp Gly Ser Thr Ser Ser Pro Gln Leu Arg Leu Pro His Thr
             130   135   140
Leu Arg Cys Ala Asn Ile Thr Ile Ile Glu His Gln Lys Cys Glu Asn
         145   150   155
Ala Tyr Pro Gly Asn Ile Thr Asp Thr Met Val Cys Ala Ser Val Gln
     160   165   170
Glu Gly Gly Lys Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val
175   180   185   190
Cys Asn Gln Ser Leu Gln Gly Ile Ile Ser Trp Gly Gln Asp Pro Cys
             195   200   205
Ala Ile Thr Arg Lys Pro Gly Val Tyr Thr Lys Val Cys Lys Tyr Val
         210   215   220
Asp Trp Ile Gln Glu Thr Met Lys Asn Asn
     225   230

```
<210> 479
<211> 151
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -21..-1

<400> 479
Met Ala Ala Ser Thr Ser Met Val Pro Val Ala Val Thr Ala Ala Val
    -20              -15              -10
Ala Pro Val Leu Ser Ile Asn Ser Asp Phe Ser Asp Leu Arg Glu Ile
-5              1              5                   10
Lys Lys Gln Leu Leu Leu Ile Ala Gly Leu Thr Arg Glu Arg Gly Leu
        15                  20                  25
Leu His Ser Ser Lys Trp Ser Ala Glu Leu Ala Phe Ser Leu Pro Ala
        30                  35                  40
Leu Pro Leu Ala Glu Leu Gln Pro Pro Pro Ile Thr Glu Glu Asp
    45                  50                  55
Ala Gln Asp Met Asp Ala Tyr Thr Leu Ala Lys Ala Tyr Phe Asp Val
60                  65                  70                  75
Lys Glu Tyr Asp Arg Ala Ala His Phe Leu His Gly Cys Asn Ala Arg
            80                  85                  90
Lys Ala Tyr Phe Leu Tyr Met Tyr Ser Arg Tyr Leu Val Arg Ala Ile
            95                  100                 105
Leu Lys Cys His Ser Ala Phe Ser Glu Thr Ser Ile Phe Arg Thr Asn
        110                 115                 120
Gly Lys Val Lys Ser Phe Lys
    125                 130


<210> 480
<211> 239
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -25..-1

<400> 480
Met Pro Arg Lys Arg Lys Cys Asp Leu Arg Ala Val Arg Val Gly Leu
-25              -20              -15                  -10
Leu Leu Gly Gly Gly Gly Val Tyr Gly Ser Arg Phe Arg Phe Thr Phe
            -5              1              5
Pro Gly Cys Arg Ala Leu Ser Pro Trp Arg Val Arg Xaa Gln Arg Arg
        10                  15                  20
Arg Cys Glu Met Ser Thr Met Phe Ala Asp Thr Leu Leu Ile Val Phe
    25                  30                  35
Ile Ser Val Cys Thr Ala Leu Leu Ala Glu Gly Ile Thr Trp Val Leu
40                  45                  50                  55
Val Tyr Arg Thr Asp Lys Tyr Lys Arg Leu Lys Ala Glu Val Glu Lys
            60                  65                  70
Gln Ser Lys Lys Leu Glu Lys Lys Lys Glu Thr Ile Thr Glu Ser Ala
        75                  80                  85
Gly Arg Gln Gln Gln Lys Lys Lys Ile Glu Arg Xaa Xaa Xaa Xaa Leu Xaa
        90                  95                  100
```

```
Asn Asn Asn Arg Asp Leu Ser Met Val Arg Met Lys Ser Met Phe Ala
    105                 110                 115
Ile Gly Phe Cys Phe Thr Ala Leu Met Gly Met Phe Asn Ser Ile Phe
120                 125                 130                 135
Asp Gly Arg Val Val Ala Lys Leu Pro Phe Thr Pro Leu Ser Xaa Xaa
            140                 145                 150
Xaa Gly Leu Ser His Arg Asn Leu Leu Gly Asp Asp Thr Thr Asp Cys
        155                 160                 165
Ser Phe Ile Phe Leu Xaa Ile Leu Cys Thr Met Ser Ile Arg Gln Asn
        170                 175                 180
Ile Gln Lys Ile Leu Gly Leu Ala Pro Ser Arg Ala Ala Thr Lys Gln
    185                 190                 195
Ala Gly Gly Phe Leu Gly Pro Pro Pro Pro Ser Gly Lys Phe Ser
200                 205                 210
```

```
<210> 481
<211> 208
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -92..-1

<400> 481
Met Arg Glu Pro Gln Lys Arg Thr Ala Thr Ile Ala Lys Xaa Xaa Ala
    -90                 -85                 -80
Xaa Glu Gly Leu Arg Asp Pro Tyr Gly Arg Leu Cys Gly Ser Glu His
    -75                 -70                 -65
Pro Arg Arg Pro Pro Glu Arg Pro Glu Glu Asp Pro Ser Thr Pro Glu
-60                 -55                 -50                 -45
Glu Ala Ser Thr Thr Pro Glu Glu Ala Ser Ser Thr Ala Gln Ala Gln
                -40                 -35                 -30
Lys Pro Ser Val Pro Arg Ser Asn Phe Gln Gly Thr Lys Lys Ser Leu
            -25                 -20                 -15
Leu Met Ser Ile Leu Ala Leu Ile Phe Ile Met Gly Asn Ser Ala Lys
        -10                 -5                  1
Glu Ala Leu Val Trp Lys Val Leu Gly Lys Leu Gly Met Gln Pro Gly
5                   10                  15                  20
Arg Xaa His Ser Ile Phe Gly Asp Pro Lys Lys Ile Val Thr Glu Xaa
                25                  30                  35
Phe Val Arg Arg Gly Tyr Leu Ile Tyr Xaa Pro Val Pro Arg Xaa Ser
            40                  45                  50
Pro Val Glu Tyr Xaa Phe Phe Trp Gly Pro Arg Ala His Val Glu Ser
        55                  60                  65
Ser Xaa Leu Lys Xaa Xaa His Phe Val Ala Arg Val Arg Asn Arg Cys
    70                  75                  80
Ser Lys Asp Trp Pro Cys Asn Tyr Asp Trp Asp Ser Asp Asp Asp Ala
85                  90                  95                  100
Glu Val Glu Ala Ile Leu Asn Ser Gly Ala Xaa Gly Tyr Ser Ala Pro
            105                 110                 115
```

```
<210> 482
<211> 86
<212> PRT
<213> Homo sapiens

<220>
```

<221> SIGNAL
<222> -39..-1

<400> 482
Met Asn Val Gly Thr Ala His Xaa Xaa Val Asn Pro Asn Thr Arg Val
            -35   -30   -25
Met Asn Ser Arg Gly Ile Trp Leu Ser Tyr Val Leu Ala Ile Gly Leu
            -20   -15   -10
Leu His Ile Val Leu Leu Ser Ile Pro Phe Val Ser Val Pro Val Val
            -5   1   5
Trp Thr Leu Thr Asn Leu Ile His Asn Met Gly Met Tyr Ile Phe Leu
10   15   20   25
His Thr Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala
            30   35   40
Arg Leu Leu Thr His Trp
            45

<210> 483
<211> 40
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -27..-1

<400> 483
Met Arg Thr Leu Phe Gly Ala Val Arg Ala Pro Phe Ser Ser Leu Thr
            -25   -20   -15
Leu Leu Leu Ile Thr Pro Ser Pro Ser Pro Leu Leu Phe Asp Arg Gly
            -10   -5   1   5
Leu Ser Leu Arg Ser Ala Met Ser
                        10

<210> 484
<211> 65
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -16..-1

<400> 484
Met Leu Gly Phe Phe Leu Phe Leu Ser Phe Val Leu Met Tyr Asp Gly
            -15   -10   -5
Leu Arg Leu Phe Gly Ile Leu Ser Thr Cys Arg Val His His Thr Met
1   5   10   15
Asn Gln Phe Leu Ile Asp Ile Ser Ser Phe Thr Ser Arg Val Lys Lys
                        20   25   30
Lys Ile Phe Leu Phe Tyr Ala Phe Xaa Gly Cys Xaa Phe Gln Ser Ala
            35   40   45
Thr

<210> 485
<211> 130

<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -55..-1

<400> 485
Met Ala Met Trp Asn Arg Pro Xaa Xaa Xaa Leu Pro Gln Gln Pro Leu
-55   -50   -45   -40
Xaa Ala Glu Pro Thr Ala Glu Gly Glu Pro His Leu Pro Thr Gly Arg
              -35   -30   -25
Xaa Xaa Thr Glu Ala Asn Arg Phe Ala Tyr Ala Ala Leu Cys Gly Ile
          -20   -15   -10
Ser Leu Ser Gln Leu Phe Pro Glu Pro Glu His Ser Ser Phe Cys Thr
      -5   1   5
Glu Phe Met Ala Gly Leu Val Xaa Trp Leu Glu Leu Ser Glu Ala Val
10   15   20   25
Leu Pro Thr Met Thr Ala Phe Ala Ser Gly Leu Gly Gly Glu Gly Xaa
            30   35   40
Xaa Cys Val Cys Ser Asn Phe Thr Glu Gly Pro His Leu Glu Gly Arg
        45   50   55
Pro Asp Gly Asp His Ser Gly Pro Ser Glu Leu Leu Thr Gln Gly Trp
      60   65   70
Ala Leu
    75

<210> 486
<211> 209
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -84..-1

<400> 486
Met Val Asn Phe Pro Gln Lys Ile Ala Gly Glu Leu Tyr Gly Pro Leu
          -80   -75   -70
Met Leu Val Phe Thr Leu Val Ala Ile Leu Leu His Gly Met Lys Thr
          -65   -60   -55
Ser Asp Thr Ile Ile Arg Glu Gly Thr Leu Met Gly Thr Ala Ile Gly
      -50   -45   -40
Thr Cys Phe Gly Tyr Trp Leu Gly Val Ser Ser Phe Ile Tyr Phe Leu
      -35   -30   -25
Ala Tyr Leu Cys Asn Ala Gln Ile Thr Met Leu Gln Met Leu Ala Leu
-20   -15   -10   -5
Leu Gly Tyr Gly Leu Phe Gly His Cys Ile Val Leu Phe Ile Thr Tyr
              1   5   10
Asn Ile His Leu Arg Ala Leu Phe Tyr Leu Phe Trp Leu Leu Val Gly
      15   20   25
Gly Leu Ser Thr Leu Arg Met Val Ala Val Leu Val Ser Arg Thr Val
      30   35   40
Gly Pro Thr Xaa Arg Xaa Leu Leu Cys Gly Thr Leu Ala Ala Leu His
45   50   55   60
Met Leu Phe Leu Leu Tyr Leu His Phe Ala Tyr His Lys Xaa Val Xaa
          65   70   75
Gly Ile Leu Asp Thr Leu Glu Gly Pro Asn Ile Pro Pro Ile Gln Arg
          80   85   90
Val Pro Arg Asp Ile Pro Ala Met Leu Pro Ala Ala Arg Leu Pro Thr

```
            95                    100                   105
Thr Val Leu Asn Ala Thr Ala Lys Ala Val Ala Val Thr Leu Gln Ser
    110                   115                   120
His
125
```

```
<210> 487
<211> 36
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -17..-1

<400> 487
Met Gly Trp Gln Arg Trp Trp Cys Phe His Leu Gln Ala Glu Ala Ser
        -15                  -10                  -5
Ala His Pro Pro Gln Gly Leu Gln Ala Gln Phe Ser Cys Cys Pro Trp
    1               5                    10                   15
Val Gly Ile Cys
```

```
<210> 488
<211> 44
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -29..-1

<400> 488
Met Met Ser Ser Glu Leu Arg Arg Asn Pro His Phe Leu Lys Ser Asn
                -25                  -20                  -15
Leu Phe Leu Gln Leu Leu Val Ser His Glu Ile Val Cys Ala Thr Glu
            -10                  -5                   1
Thr Val Thr Thr Asn Phe Leu Arg His Glu Lys Ala
    5                   10                   15
```

```
<210> 489
<211> 163
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -52..-1

<400> 489
Met Glu His Tyr Arg Lys Ala Gly Ser Val Glu Leu Pro Ala Pro Ser
        -50                  -45                  -40
Pro Met Pro Gln Leu Pro Pro Asp Thr Leu Glu Met Arg Val Arg Asp
    -35                  -30                  -25
Gly Ser Lys Ile Arg Asn Leu Leu Gly Leu Ala Leu Gly Arg Leu Glu
-20                  -15                  -10                  -5
Gly Gly Ser Ala Arg His Val Val Phe Ser Gly Ser Gly Arg Ala Ala
```

```
            1               5                 10
Gly Lys Ala Val Ser Cys Ala Glu Ile Val Lys Arg Arg Val Pro Gly
        15              20                25
Leu His Gln Leu Thr Lys Leu Xaa Phe Leu Gln Thr Glu Asp Ser Trp
    30              35                40
Val Pro Xaa Ser Pro Asp Thr Gly Leu Xaa Pro Leu Thr Val Arg Arg
45              50                55                60
His Val Pro Ala Xaa Trp Val Leu Leu Xaa Arg Asp Pro Leu Asp Pro
            65                  70                75
Asn Glu Cys Gly Tyr Gln Pro Pro Gly Ala Pro Pro Gly Leu Gly Ser
            80              85                90
Met Pro Ser Ser Ser Cys Gly Pro Arg Ser Xaa Lys Arg Ala Xaa Xaa
        95              100               105
Thr Arg Ser
    110
```

```
<210> 490
<211> 64
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -47..-1

<400> 490
Met His Gly Phe Glu Ile Ile Ser Leu Lys Glu Glu Ser Pro Leu Gly
        -45             -40               -35
Lys Val Ser Gln Gly Pro Leu Phe Asn Val Thr Ser Gly Ser Ser Ser
    -30             -25               -20
Pro Val Thr Trp Leu Gly Leu Leu Ser Phe Gln Asn Leu His Cys Phe
-15             -10               -5                 1
Pro Asp Leu Pro Thr Glu Met Pro Leu Xaa Ala Lys Gly Xaa Asn Thr
            5               10                15
```

```
<210> 491
<211> 218
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -50..-1

<400> 491
Met His His Gly Leu Thr Pro Leu Leu Leu Gly Val His Glu Gln Lys
-50             -45             -40               -35
Gln Gln Val Val Lys Phe Leu Ile Lys Lys Lys Ala Asn Leu Asn Ala
            -30             -25               -20
Leu Asp Arg Tyr Gly Arg Thr Ala Leu Ile Leu Ala Val Cys Cys Gly
            -15             -10               -5
Ser Ala Ser Ile Val Ser Leu Leu Leu Glu Gln Asn Ile Asp Val Ser
        1               5               10
Ser Gln Asp Leu Ser Gly Gln Thr Ala Lys Lys Tyr Ala Val Ser Ser
15              20                25                30
Arg His Asn Val Ile Cys Gln Leu Leu Ser Asp Tyr Lys Xaa Lys Gln
            35              40                45
Xaa Leu Lys Val Ser Ser Glu Asn Ser Asn Pro Xaa Gln Asp Leu Lys
```

```
                 50                      55                      60
Leu Thr Ser Glu Glu Glu Ser Gln Arg Leu Lys Gly Ser Glu Asn Ser
            65                      70                      75
Gln Pro Glu Glu Met Ser Gln Glu Pro Glu Ile Asn Xaa Gly Gly Asp
        80                      85                      90
Arg Lys Val Glu Xaa Xaa Met Lys Lys His Gly Ser Xaa His Met Gly
95                      100                     105                     110
Phe Pro Xaa Asn Leu Xaa Asn Gly Ala Thr Ala Asp Asn Gly Asp Asp
                115                     120                     125
Gly Leu Ile Pro Pro Xaa Lys Xaa Xaa Thr Pro Glu Ser Xaa Gln Phe
            130                     135                     140
Pro Asp Thr Glu Asn Glu Gln Tyr His Arg Asp Phe Ser Gly His Pro
            145                     150                     155
Xaa Phe Pro Thr Thr Leu Pro Ile Lys Gln
    160                     165
```

```
<210> 492
<211> 216
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -15..-1

<400> 492
Met Val Cys Val Leu Val Leu Ala Ala Ala Ala Gly Ala Val Ala Val
-15                     -10                     -5                      1
Phe Leu Ile Leu Arg Ile Trp Val Val Leu Arg Ser Met Asp Val Thr
            5                       10                      15
Pro Arg Glu Ser Leu Ser Ile Leu Val Val Ala Gly Ser Gly Gly His
        20                      25                      30
Thr Thr Glu Ile Leu Arg Leu Leu Gly Ser Leu Ser Asn Ala Tyr Ser
        35                      40                      45
Pro Arg His Tyr Val Ile Ala Asp Thr Asp Glu Met Ser Ala Asn Lys
50                      55                      60                      65
Ile Asn Ser Phe Glu Leu Xaa Arg Xaa Asp Arg Xaa Pro Ser Asn Met
                70                      75                      80
Xaa Thr Lys Tyr Tyr Ile His Arg Ile Pro Xaa Ser Arg Glu Val Gln
            85                      90                      95
Gln Ser Trp Pro Ser Thr Val Xaa Thr Thr Leu His Ser Met Trp Leu
        100                     105                     110
Ser Xaa Pro Leu Ile His Arg Val Lys Pro Xaa Leu Val Leu Cys Asn
    115                     120                     125
Gly Pro Gly Thr Cys Val Pro Ile Cys Val Ser Ala Leu Leu Leu Gly
130                     135                     140                     145
Ile Leu Gly Ile Lys Lys Val Ile Ile Val Tyr Val Glu Ser Ile Cys
                150                     155                     160
Arg Val Lys Thr Leu Ser Met Ser Gly Lys Ile Leu Phe His Leu Ser
            165                     170                     175
Asn Tyr Phe Ile Val Gln Trp Pro Ala Leu Lys Glu Lys Tyr Pro Lys
            180                     185                     190
Ser Val Tyr Leu Gly Arg Ile Val
    195                     200
```

```
<210> 493
<211> 134
<212> PRT
```

<213> Homo sapiens

<220>
<221> SIGNAL
<222> -19..-1

<400> 493
Met Pro Leu Gly Ala Arg Ile Leu Phe His Gly Val Phe Tyr Ala Gly
            -15   -10   -5
Gly Phe Ala Ile Val Tyr Tyr Leu Ile Gln Lys Phe His Ser Arg Thr
            1   5   10
Leu Tyr Tyr Lys Leu Ala Val Glu Gln Leu Gln Xaa His Pro Glu Ala
    15   20   25
Gln Glu Ala Leu Gly Pro Pro Leu Asn Ile His Tyr Leu Lys Leu Ile
30   .35.   40   45
Asp Arg Glu Asn Phe Val Asp Ile Val Xaa Ala Lys Leu Lys Ile Pro
            50   55   60
Val Ser Gly Ser Lys Ser Glu Gly Leu Leu Tyr Val His Ser Ser Arg
            65   70   75
Gly Gly Pro Phe Gln Arg Trp His Leu Asp Glu Val Phe Leu Glu Leu
            80   85   90
Lys Asp Gly Gln Gln Ile Pro Val Phe Lys Leu Ser Gly Glu Asn Gly
    95   100   105
Asp Glu Val Lys Lys Glu
110   115

<210> 494
<211> 85
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -16..-1

<400> 494
Met Ala Val Thr Ala Leu Ala Ala Xaa Thr Trp Leu Gly Val Trp Gly
    -15   -10   -5
Val Arg Thr Met Gln Ala Arg Gly Phe Gly Ser Asp Gln Ser Glu Asn
1   5   10   15
Val Asp Arg Gly Ala Gly Ser Ile Arg Glu Ala Gly Gly Ala Phe Gly
            20   25   30
Lys Arg Glu Gln Ala Glu Glu Glu Arg Tyr Phe Arg Ala Gln Ser Thr
            35   40   45
Glu Gln Leu Ala Xaa Leu Lys Lys Xaa His Glu Glu Glu Ile Val His
    50   55   60
His Arg Glu Gly Asp
65

<210> 495
<211> 292
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -29..-1

```
<400> 495
Met His Gly Leu Leu His Tyr Leu Phe His Thr Arg Asn His Thr Phe
            -25                   -20                   -15
Ile Val Leu His Leu Val Leu Gln Gly Met Val Tyr Thr Glu Tyr Thr
            -10                    -5                    1
Trp Glu Val Phe Gly Tyr Cys Gln Glu Leu Glu Leu Ser Leu His Tyr
 5                   10                   15
Leu Leu Leu Pro Tyr Leu Leu Leu Gly Val Asn Leu Phe Phe Phe Thr
20                   25                   30                   35
Leu Thr Cys Gly Thr Asn Pro Gly Ile Ile Thr Lys Ala Asn Glu Leu
            40                   45                   50
Leu Phe Leu His Val Tyr Glu Phe Asp Glu Xaa Met Phe Pro Lys Asn
            55                   60                   65
Val Arg Cys Ser Thr Cys Asp Leu Arg Lys Pro Ala Arg Ser Xaa His
            70                   75                   80
Cys Xaa Val Cys Asn Trp Cys Val His Arg Phe Xaa His His Cys Val
            85                   90                   95
Trp Val Asn Asn Cys Ile Gly Ala Trp Asn Ile Arg Xaa Phe Leu Ile
100                  105                  110                  115
Tyr Val Leu Thr Leu Thr Ala Ser Ala Ala Thr Val Ala Ile Val Ser
            120                  125                  -    130
Thr Thr Phe Leu Val His Leu Val Val Met Ser Asp Leu Tyr Gln Glu
            135                  140                  145
Thr Tyr Ile Asp Asp Leu Gly His Leu His Val Met Asp Thr Val Phe
            150                  155                  160
Leu Ile Gln Tyr Leu Phe Leu Thr Phe Pro Arg Ile Val Phe Met Leu
            165                  170                  175
Gly Phe Val Val Val Leu Xaa Phe Leu Leu Gly Gly Tyr Leu Leu Phe
180                  185                  190                  195
Val Leu Tyr Leu Ala Ala Thr Asn Gln Thr Thr Asn Glu Trp Tyr Arg
            200                  205                  210
Xaa Asp Trp Ala Trp Cys Gln Arg Cys Pro Leu Val Ala Trp Pro Pro
            215                  220                  225
Ser Ala Glu Pro Gln Val His Arg Asn Ile His Ser His Gly Leu Arg
            230                  235                  240
Xaa Asn Leu Gln Glu Ile Phe Leu Pro Ala Phe Pro Cys His Glu Arg
            245                  250                  255
Lys Lys Gln Glu
260


<210> 496
<211> 122
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -56..-1

<400> 496
Met Thr Gly Phe Leu Leu Pro Pro Ala Ser Arg Gly Thr Arg Arg Ser
    -55                   -50                   -45
Cys Ser Arg Ser Arg Lys Arg Gln Thr Arg Arg Arg Arg Asn Pro Ser
-40                   -35                   -30                   -25
Ser Phe Val Ala Ser Cys Pro Thr Leu Leu Pro Phe Ala Cys Val Pro
            -20                   -15                   -10
Gly Ala Ser Pro Thr Thr Leu Ala Phe Pro Pro Val Xaa Leu Thr Gly
             -5                    1                    5
Pro Xaa Thr Asp Gly Ile Pro Phe Ala Leu Xaa Ser Ala Ala Gly Pro
    10                   15                   20
```

```
Phe Cys Ala Ser Phe Pro Ser Gly Xaa Leu Ser Pro Pro Gly Pro Leu
25                  30              35              40
Pro Gly Val Arg Gly Leu Pro Leu Pro Ser Val Phe Tyr Ser Cys Gly
            45              50              55
Ala His Pro Lys Val Leu Lys Val Ala Leu
        60              65
```

<210> 497
<211> 59
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -28..-1

<400> 497
```
Met Leu Xaa Leu Ser Arg Ala Thr Lys Xaa Gly Arg Ala Arg Trp Leu
        -25             -20             -15
Met Pro Val Ile Pro Ala Leu Gln Glu Ala Xaa Ala Gly Gly Ser Arg
        -10             -5              1
Gly Gln Glu Phe Glu Thr Ser Leu Ala Asn Met Glu Thr Glu Ala Gly
5                   10              15              20
Glu Leu Leu Lys Pro Arg Arg Arg Arg Leu Gln
                25              30
```

<210> 498
<211> 99
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -13..-1

<400> 498
```
Met His Leu Leu Ser Asn Trp Ala Asn Pro Ala Ser Ser Arg Arg Pro
            -10             -5              1
Ser Met Ala Ala Ser Gly Thr Ser Trp Ile Ser Ser Thr Leu Ala His
    5               10              15
Ser Leu Ser Leu Arg Asp Val Ser Glu Arg Leu Cys Ser Cys Trp Arg
20              25              30              35
Thr Ile Ser Met Gly Pro Cys Ala Arg Gly Ser Pro Met Asn Ser Ser
            40              45              50
Gly Val His Arg Lys Ser Ser Arg Leu Phe Tyr Ile Arg Thr Pro Met
        55              60              65
Arg Arg Ser Ser Cys His Leu Glu Cys Xaa Val Ile Phe Leu Leu Gly
        70              75              80
Arg Gln Leu
    85
```

<210> 499
<211> 99
<212> PRT
<213> Homo sapiens

```
<220>
<221> SIGNAL
<222> -13..-1

<400> 499
Met His Leu Leu Ser Asn Trp Ala Asn Pro Ala Ser Ser Arg Arg Pro
        -10                 -5                  1
Ser Met Ala Ala Ser Gly Thr Ser Trp Ile Ser Ser Thr Leu Ala His
    5               10                  15
Ser Leu Ser Leu Arg Asp Val Ser Glu Arg Leu Cys Ser Cys Trp Arg
20              25                  30                      35
Thr Ile Ser Met Gly Pro Cys Ala Arg Gly Ser Pro Met Asn Ser Ser
            40                  45                      50
Gly Val His Arg Lys Ser Ser Arg Leu Phe Tyr Ile Arg Thr Pro Met
        55                  60                  65
Arg Arg Ser Ser Cys His Leu Xaa Cys Gln Val Ile Phe Leu Leu Gly
        70                  75                  80
Arg Gln Leu
    85
```

```
<210> 500
<211> 108
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -25..-1

<400> 500
Met Ser Leu Thr Ser Ser Ser Ser Val Arg Val Glu Trp Ile Ala Ala
-25             -20                 -15                     -10
Val Thr Ile Ala Ala Gly Thr Ala Ala Ile Gly Tyr Leu Ala Tyr Lys
            -5                  1                   5
Arg Phe Tyr Val Lys Asp His Arg Asn Lys Ala Met Ile Asn Leu His
        10                  15                  20
Ile Gln Lys Asp Asn Pro Lys Ile Val His Ala Phe Asp Met Glu Asp
    25                  30                  35
Leu Gly Asp Lys Ala Val Tyr Cys Arg Cys Trp Arg Ser Lys Lys Phe
40                  45                  50                      55
Pro Phe Cys Asp Gly Ala His Thr Lys His Asn Glu Glu Thr Gly Asp
            60                  65                      70
Asn Val Gly Pro Leu Ile Ile Lys Lys Lys Glu Thr
            75                  80
```

```
<210> 501
<211> 183
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -15..-1

<400> 501
Met Glu Ala Met Trp Leu Leu Cys Val Ala Leu Ala Val Leu Ala Trp
-15             -10                 -5                      1
Gly Phe Leu Trp Val Trp Asp Ser Ser Glu Arg Met Lys Ser Arg Glu
```

```
                    5                      10                      15
Gln Gly Arg Arg Leu Gly Ala Glu Ser Arg Thr Leu Leu Val Ile Ala
            20                      25                      30
His Pro Asp Asp Glu Ala Met Phe Phe Ala Pro Thr Val Leu Gly Leu
        35                      40                      45
Ala Arg Leu Arg His Trp Val Tyr Leu Leu Cys Phe Ser Ala Gly Asn
50                      55                      60                      65
Tyr Tyr Asn Gln Gly Glu Thr Arg Lys Lys Glu Leu Leu Gln Ser Cys
                    70                      75                      80
Asp Val Leu Gly Ile Pro Leu Ser Ser Val Met Ile Ile Asp Asn Arg
                85                      90                      95
Asp Phe Pro Xaa Asp Pro Gly Met Gln Trp Asp Thr Xaa His Val Ala
            100                     105                     110
Xaa Val Leu Leu Gln His Ile Glu Val Asn Gly Ile Asn Leu Val Val
    115                     120                     125
Thr Phe Asp Ala Gly Gly Xaa Ser Gly His Ser Asn His Ile Ala Leu
130                     135                     140                     145
Tyr Ala Ala Val Arg Lys Leu Glu Gly Gln Ile Cys Lys Pro Cys Gly
                150                     155                     160
Thr Gly Gln Asp Phe Lys Glu
                165
```

```
<210> 502
<211> 98
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -15..-1

<400> 502
Met Glu Ala Met Trp Leu Leu Cys Val Ala Leu Ala Val Leu Ala Trp
-15                     -10                     -5                      1
Gly Phe Leu Trp Val Trp Asp Ser Ser Glu Arg Met Lys Ser Arg Glu
            5                       10                      15
Gln Gly Xaa Arg Leu Gly Ala Glu Ser Arg Thr Leu Leu Val Ile Ala
            20                      25                      30
His Pro Asp Asp Glu Ala Met Phe Phe Ala Pro Thr Val Leu Gly Leu
        35                      40                      45
Ala Arg Leu Arg His Trp Val Tyr Leu Leu Cys Phe Ser Ala Val Phe
50                      55                      60                      65
Arg Arg Glu Leu Ser Glu Tyr Thr Glu Xaa Leu Thr Ser Glu Pro Leu
                70                      75                      80
Xaa Ala
```

```
<210> 503
<211> 183
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -57..-1

<400> 503
Met Asp Val Thr Gly Asp Glu Glu Glu Glu Ile Lys Gln Glu Ile Asn
        -55                     -50                     -45
```

```
Met Leu Lys Lys Tyr Ser His His Arg Asn Ile Ala Thr Tyr Tyr Gly
    -40                 -35                 -30
Ala Phe Ile Lys Lys Asn Pro Pro Gly Met Asp Asp Gln Leu Trp Leu
-25                 -20                 -15                     -10
Val Met Glu Phe Cys Gly Ala Gly Ser Val Thr Asp Leu Ile Lys Asn
            -5                  1                   5
Thr Lys Gly Asn Thr Leu Lys Glu Glu Trp Ile Ala Tyr Ile Cys Xaa
        10                  15                  20
Glu Ile Leu Arg Gly Leu Xaa His Leu His Gln His Lys Val Ile His
        25                  30                  35
Arg Xaa Ile Lys Gly Gln Asn Val Leu Leu Thr Glu Asn Ala Glu Val
40                  45                  50                      55
Lys Leu Val Asp Phe Gly Xaa Xaa Ala Gln Leu Asp Arg Thr Val Gly
                60                  65                  70
Arg Xaa Asn Thr Phe Ile Gly Thr Pro Tyr Trp Met Ala Pro Xaa Val
            75                  80                  85
Ile Ala Cys Asp Glu Asn Pro Xaa Ala Thr Tyr Asp Phe Lys Xaa Asp
        90                  95                  100
Leu Trp Ser Leu Gly Ile Thr Ala Ile Glu Met Ala Glu Gly Leu Pro
        105                 110                 115
Leu Ser Val Thr Cys Thr Pro
120                 125
```

```
<210> 504
<211> 140
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -14..-1

<400> 504
Met Phe Leu Thr Ala Leu Leu Trp Arg Gly Arg Ile Pro Gly Arg Gln
            -10                 -5                  1
Trp Ile Gly Lys His Arg Arg Pro Arg Phe Val Ser Leu Arg Ala Lys
        5                   10                  15
Gln Asn Met Ile Arg Arg Leu Glu Ile Glu Ala Glu Asn His Tyr Trp
        20                  25                  30
Leu Ser Met Pro Tyr Met Thr Arg Glu Gln Glu Arg Gly His Ala Ala
35                  40                  45                      50
Leu Arg Arg Arg Glu Ala Phe Glu Ala Ile Lys Ala Ala Ala Thr Ser
                55                  60                  65
Lys Phe Pro Pro His Arg Phe Ile Ala Asp Gln Leu Asp His Leu Asn
            70                  75                  80
Xaa His Gln Glu Met Val Leu Ile Leu Ser Arg His Pro Trp Ile Leu
        85                  90                  95
Trp Ile Thr Glu Leu Thr Ile Phe Thr Trp Ser Gly Leu Lys Asn Cys
        100                 105                 110
Ser Leu Cys Glu Asn Glu Leu Trp Thr Ser Leu Tyr
115                 120                 125
```

```
<210> 505
<211> 59
<212> PRT
<213> Homo sapiens

<220>
```

<221> SIGNAL
<222> -14..-1

<400> 505
Met Ala Ala Leu Val Thr Val Leu Phe Thr Gly Val Arg Arg Leu His
             -10   -5   1
Cys Ser Ala Xaa Leu Gly Arg Ala Ala Ser Gly Xaa Tyr Ser Arg Asn
     5   10   15
Trp Leu Pro Thr Pro Pro Ala Thr Gly Pro Leu Pro Ser Ser Gln Thr
  20   25   30
Gly His Met Arg Met Ala Ala Leu Leu Pro Gln
35   40   45

<210> 506
<211> 101
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -36..-1

<400> 506
Met Gly Pro Tyr Asn Val Ala Val Pro Ser Asp Val Ser His Ala Arg
     -35   -30   -25
Phe Tyr Phe Leu Phe His Arg Pro Leu Arg Leu Leu Asn Leu Leu Ile
-20   -15   -10   -5
Leu Ile Glu Gly Ser Val Val Phe Tyr Gln Leu Tyr Ser Leu Leu Arg
          1   5   10
Ser Glu Lys Trp Asn His Thr Leu Ser Met Ala Leu Ile Leu Phe Cys
     15   20   25
Asn Tyr Tyr Val Leu Phe Lys Leu Leu Arg Asp Arg Xaa Xaa Leu Gly
    30   35   40
Arg Ala Tyr Ser Tyr Pro Leu Asn Ser Tyr Glu Leu Lys Ala Asn Xaa
45   50   55   60
Ala Ala Ser Xaa Gln
        65

<210> 507
<211> 341
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -55..-1

<400> 507
Met Arg Lys Val Val Leu Ile Thr Gly Ala Ser Ser Gly Ile Gly Leu
-55   -50   -45   -40
Ala Leu Cys Lys Arg Leu Leu Ala Glu Asp Asp Glu Leu His Leu Cys
        -35   -30   -25
Leu Ala Cys Arg Asn Met Ser Lys Ala Glu Ala Val Cys Ala Ala Leu
        -20   -15   -10
Leu Ala Ser His Pro Thr Ala Glu Val Thr Ile Val Gln Val Asp Val
       -5   1   5
Ser Asn Leu Gln Ser Phe Phe Arg Ala Ser Lys Glu Leu Lys Gln Arg
10   15   20   25

```
Phe Gln Arg Leu Asp Cys Ile Tyr Leu Asn Ala Gly Ile Met Pro Asn
            30                  35                  40
Pro Gln Leu Asn Ile Lys Ala Leu Phe Phe Gly Leu Phe Ser Arg Lys
            45                  50                  55
Val Ile His Met Phe Ser Thr Ala Glu Gly Leu Leu Thr Gln Gly Asp
            60                  65                  70
Lys Ile Thr Ala Asp Gly Leu Gln Glu Val Phe Glu Thr Asn Val Phe
        75                  80                  85
Gly His Phe Ile Leu Ile Arg Glu Leu Glu Pro Leu Leu Cys His Ser
90                  95                  100                 105
Asp Asn Pro Ser Gln Leu Ile Trp Thr Ser Ser Arg Ser Ala Arg Lys
                110                 115                 120
Ser Asn Phe Ser Leu Glu Asp Phe Gln His Ser Lys Gly Lys Glu Pro
            125                 130                 135
Tyr Ser Ser Ser Lys Tyr Ala Thr Asp Leu Leu Ser Val Ala Leu Asn
        140                 145                 150
Arg Asn Phe Asn Gln Gln Gly Leu Tyr Ser Asn Val Ala Cys Pro Gly
    155                 160                 165
Thr Ala Leu Thr Asn Leu Thr Tyr Gly Ile Leu Pro Pro Phe Ile Trp
170                 175                 180                 185
Thr Leu Leu Met Pro Ala Ile Leu Leu Leu Arg Phe Phe Ala Asn Ala
                190                 195                 200
Phe Thr Leu Thr Pro Tyr Asn Gly Thr Glu Ala Leu Val Trp Leu Phe
            205                 210                 215
His Gln Lys Pro Glu Ser Leu Asn Pro Leu Ile Lys Tyr Leu Ser Ala
        220                 225                 230
Thr Thr Gly Phe Gly Arg Asn Tyr Ile Met Thr Gln Lys Met Asp Leu
    235                 240                 245
Asp Glu Asp Thr Ala Glu Lys Phe Tyr Gln Lys Leu Leu Glu Leu Glu
250                 255                 260                 265
Lys His Ile Arg Val Thr Ile Gln Lys Thr Asp Asn Gln Ala Arg Leu
                270                 275                 280
Ser Gly Ser Cys Leu
            285




<210> 508
<211> 108
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -42..-1

<400> 508
Met His Ile Leu Gln Leu Leu Thr Thr Val Asp Asp Gly Ile Gln Ala
        -40                 -35                 -30
Ile Val His Cys Pro Asp Thr Gly Lys Asp Ile Trp Asn Leu Leu Phe
    -25                 -20                 -15
Asp Leu Val Cys His Glu Phe Cys Gln Ser Asp Asp Pro Ala Ile Ile
-10                 -5                  1               5
Leu Gln Xaa Gln Lys Thr Val Leu Ala Ser Val Phe Ser Val Leu Ser
            10                  15                  20
Ala Ile Tyr Ala Ser Gln Thr Glu Gln Xaa Tyr Leu Lys Ile Xaa Lys
        25                  30                  35
Gly Asp Gly Gly Ser Gly Ser Lys Gly Arg Pro Xaa Xaa Gln Thr Glu
        40                  45                  50
Xaa Phe Leu Cys Ile Ser Lys Pro Ser Ser Phe Leu
55                  60                  65
```

<210> 509
<211> 80
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -26..-1

<400> 509
Met Glu Glu Ile Ser Ser Pro Leu Val Glu Phe Val Lys Val Leu Cys
    -25   -20   -15
Thr Asn Gln Val Leu Ile Thr Ala Arg Ala Val Pro Thr Lys Lys Ala
-10   -5   1   5
Ser Val Arg Cys Val Glu Lys Arg Phe Trp Ile Pro Lys Thr Thr Ser
              10   15   20
Lys His Leu Ser Arg Cys Ile Asp Gly Ile Ser Gly Phe Leu Asn Asp
          25   30   35
Phe Thr Phe Cys Leu Glu Phe Ser Arg His Arg Cys Gln Leu Thr Glu
    40   45   50

<210> 510
<211> 158
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -44..-1

<400> 510
Met Ala Gly Phe Leu Asp Asn Phe Arg Trp Pro Glu Cys Glu Cys Ile
              -40   -35   -30
Asp Trp Ser Glu Arg Arg Asn Ala Val Ala Ser Val Val Ala Gly Ile
              -25   -20   -15
Leu Phe Phe Thr Gly Trp Trp Ile Met Ile Asp Ala Ala Val Val Tyr
              -10   -5   1
Pro Lys Pro Glu Gln Leu Asn His Ala Phe His Thr Cys Gly Val Phe
5   10   15   20
Ser Thr Leu Ala Phe Phe Met Ile Asn Ala Val Ser Asn Ala Gln Val
              25   30   35
Arg Gly Asp Ser Tyr Glu Ser Gly Cys Leu Gly Arg Thr Gly Ala Arg
              40   45   50
Val Trp Leu Phe Ile Gly Phe Met Leu Met Phe Gly Ser Leu Ile Ala
              55   60   65
Ser Met Trp Ile Leu Phe Gly Ala Tyr Val Thr Gln Asn Thr Asp Val
    70   75   80
Tyr Pro Gly Leu Ala Val Phe Phe Gln Asn Ala Leu Ile Phe Phe Ser
85   90   95   100
Thr Leu Ile Tyr Lys Phe Gly Arg Thr Glu Glu Leu Trp Thr
              105   110

<210> 511
<211> 130
<212> PRT
<213> Homo sapiens

```
<220>
<221> SIGNAL
<222> -28..-1

<400> 511
Met Asn Trp Glu Leu Leu Leu Trp Leu Leu Val Leu Cys Ala Leu Leu
            -25                 -20                 -15
Leu Leu Leu Val Gln Leu Leu Arg Phe Leu Arg Ala Asp Gly Asp Leu
        -10                 -5                   1
Thr Leu Leu Trp Ala Glu Trp Gln Gly Arg Arg Pro Glu Trp Glu Leu
5               10                  15                      20
Thr Asp Met Val Val Trp Val Thr Gly Ala Ser Ser Gly Ile Gly Glu
                25                  30                  35
Glu Leu Ala Tyr Gln Leu Ser Lys Leu Gly Val Ser Leu Val Leu Ser
            40                  45                  50
Ala Arg Arg Val His Glu Leu Glu Arg Val Lys Arg Arg Cys Leu Glu
        55                  60                  65
Asn Gly Asn Leu Lys Glu Lys Asp Ile Leu Val Leu Pro Leu Asp Leu
    70                  75                  80
Thr Asp Thr Gly Ser His Glu Ser Gly Tyr Gln Ser Cys Ser Pro Gly
85                  90                  95                  100
Ile Trp


<210> 512
<211> 199
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -62..-1

<400> 512
Met Ser Gln Arg Ser Leu Cys Met Asp Thr Ser Leu Asp Val Tyr Arg
        -60                 -55                 -50
Xaa Leu Ile Glu Leu Asn Tyr Leu Gly Thr Val Ser Leu Thr Lys Cys
    -45                 -40                 -35
Val Leu Pro His Met Ile Glu Arg Lys Gln Gly Lys Ile Val Thr Val
-30                 -25                 -20                 -15
Asn Ser Ile Leu Gly Ile Ile Ser Val Pro Leu Ser Ile Gly Tyr Cys
                -10                 -5                   1
Ala Ser Lys His Ala Leu Arg Gly Phe Phe Asn Gly Leu Arg Thr Glu
        5                   10                  15
Leu Ala Thr Tyr Pro Gly Ile Ile Val Ser Asn Ile Cys Pro Gly Pro
    20                  25                  30
Val Gln Ser Asn Ile Val Glu Asn Ser Leu Ala Gly Glu Val Thr Lys
35                  40                  45                  50
Thr Ile Gly Asn Asn Gly Asn Gln Ser His Lys Met Thr Thr Ser Arg
                55                  60                  65
Cys Val Arg Leu Met Leu Ile Ser Met Ala Asn Asp Leu Lys Glu Val
            70                  75                  80
Trp Ile Ser Glu Gln Pro Phe Leu Leu Val Thr Tyr Leu Trp Gln Tyr
        85                  90                  95
Met Pro Thr Trp Ala Trp Trp Ile Thr Asn Lys Met Gly Lys Lys Arg
    100                 105                 110
Ile Glu Asn Phe Lys Ser Gly Val Asp Ala Xaa Ser Ser Tyr Phe Lys
115                 120                 125                 130
Ile Phe Lys Thr Lys His Asp
                135
```

<210> 513
<211> 180
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -25..-1

<400> 513
Met Asn Thr Val Leu Ser Arg Ala Asn Ser Leu Phe Ala Phe Ser Leu
-25   -20   -15   -10
Ser Val Met Ala Ala Leu Thr Phe Gly Cys Phe Ile Xaa Thr Ala Phe
                 -5   1   5
Lys Asp Arg Ser Val Pro Val Arg Leu His Val Ser Arg Ile Met Leu
            10   15   20
Lys Asn Val Glu Asp Phe Thr Gly Pro Arg Glu Arg Ser Asp Leu Gly
     25   30   35
Phe Ile Thr Phe Asp Ile Thr Ala Asp Leu Glu Asn Ile Phe Asp Trp
40   45   50   55
Asn Val Lys Gln Leu Phe Leu Tyr Leu Ser Ala Glu Tyr Ser Thr Lys
                60   65   70
Asn Asn Ala Leu Asn Gln Xaa Val Leu Trp Asp Lys Ile Val Leu Arg
                75   80   85
Gly Asp Asn Pro Lys Leu Leu Leu Lys Asp Met Lys Thr Lys Tyr Phe
            90   95   100
Phe Phe Asp Asp Gly Asn Gly Leu Xaa Gly Asn Arg Asn Val Thr Leu
     105   110   115
Thr Leu Ser Trp Asn Val Val Pro Asn Ala Gly Ile Leu Pro Leu Val
120   125   130   135
Thr Gly Ser Gly His Val Ser Val Pro Phe Pro Asp Thr Tyr Glu Ile
                140   145   150
Thr Lys Ser Tyr
            155

<210> 514
<211> 120
<212> PRT
<213> Bos taurus

<400> 514
Met Met Thr Gly Arg Gln Gly Arg Ala Thr Phe Gln Phe Leu Pro Asp
1   5   10   15
Glu Ala Arg Ser Leu Pro Pro Pro Lys Leu Thr Asp Pro Arg Leu Ala
            20   25   30
Phe Val Gly Phe Leu Gly Tyr Cys Ser Gly Leu Ile Asp Asn Ala Ile
            35   40   45
Arg Arg Arg Pro Val Leu Leu Ala Gly Leu His Arg Gln Leu Leu Tyr
     50   55   60
Ile Thr Ser Phe Val Phe Val Gly Tyr Tyr Leu Leu Lys Arg Gln Asp
65   70   75   80
Tyr Met Tyr Ala Val Arg Asp His Asp Met Phe Ser Tyr Ile Lys Ser
                85   90   95
His Pro Glu Asp Phe Pro Glu Lys Asp Lys Lys Thr Tyr Gly Glu Val
            100   105   110
Phe Glu Glu Phe His Pro Val Arg
            115   120

<210> 515
<211> 1082
<212> DNA
<213> Homo sapiens

<400> 515
```
gatcccagac ctcggcttgc agtagtgtta gactgaagat aaagtaagtg ctgtttgggc      60
taacaggatc tcctcttgca gtctgcagcc caggacgctg attccagcag cgccttaccg     120
cgcagcccga agattcacta tggtgaaaat cgccttcaat acccctaccg ccgtgcaaaa     180
ggaggaggcg cggcaagacg tggaggccct cctgagccgc acggtcagaa ctcagatact     240
gaccggcaag gagctccgag ttgccaccca ggaaaaagag ggctcctctg ggagatgtat     300
gcttactctc ttaggccttt cattcatctt ggcaggactt attgttggtg gagcctgcat     360
ttacaagtac ttcatgccca agagcaccat ttaccgtgga gagatgtgct tttttgattc     420
tgaggatcct gcaaattccc ttcgtggagg agagctcaac ttcctgcctg tgactgagga     480
ggctgacatt cgtgaggatg acaacattgc aatcattgat gtgcctgtcc ccagtttctc     540
tgatagtgac cctgcagcaa ttattcatga ctttgaaaag ggaatgactg cttacctgga     600
cttgttgctg gggaactgct atctgatgcc cctcaatact tctattgtta tgcctccaaa     660
aaatctggta gagctctttg gcaaactggc gagtggcaga tatctgcctc aaacttatgt     720
ggttcgagaa gacctagttg ctgtggagga aattcgtgat gttagtaacc ttggcatctt     780
tatttaccaa ctttgcaata acagaaagtc cttccgcctt cgtcgcagag acctcttgct     840
gggtttcaac aaacgtgcca ttgataaatg ctggaagatt agacacttcc ccaacgaatt     900
tattgttgag accaagatct gtcaagagta agaggcaaca gatagagtgt ccttggtaat     960
aagaagtcag agatttacaa tatgacttta acattaaggt ttatgggata tcaagatat    1020
ttactcatgc atttactcta ttgcttatgc cgtaaaaaaa aaaaaaaaaa aaaaaaaaa    1080
aa                                                                   1082
```


<210> 516
<211> 559
<212> DNA
<213> Homo sapiens

<400> 516
```
ctgctccagc gctgacgccg agccatggcg gacgaggagc ttgaggcgct gaggagacag      60
aggctggccg agctgcaggc caaacacggg gatcctggtg atgcggccca acaggaagca     120
aagcacaggg aagcagaaat gagaaacagt atcttagccc aagttctgga tcagtcggcc     180
cgggccaggt taagtaactt agcacttgta aagcctgaaa aaactaaagc agtagagaat     240
taccttatac agatggcaag atatggacaa ctaagtgaga aggtatcaga acaaggttta     300
atagaaatcc ttaaaaaagt aagccaacaa acagaaaaga caacaacagt gaaattcaac     360
agaagaaaag taatggactc tgatgaagat gacgattatt gaactacaag tgctcacaga     420
ctagaactta acggaacaag tctaggacag aagttaagat ctgattattt actttgttta     480
ttgtctatat gcctttttaaa aaaataaact tgttatgcaa aaaaaaaaaa aaaaaaaaaa     540
aaaaaaaaaa aaaaaaaaa                                                  559
```


<210> 517
<211> 110
<212> PRT
<213> Homo sapiens

<400> 517
```
Met Phe Cys Pro Leu Lys Leu Ile Leu Leu Pro Val Leu Leu Asp Tyr
1               5                   10                  15
Ser Leu Gly Leu Asn Asp Leu Asn Val Ser Pro Pro Glu Leu Thr Val
            20                  25                  30
His Val Gly Asp Ser Ala Leu Met Gly Cys Val Phe Gln Ser Thr Glu
        35                  40                  45
```

```
Asp Lys Cys Ile Phe Lys Ile Asp Trp Thr Leu Ser Pro Gly Glu His
    50                  55                  60
Ala Lys Asp Glu Tyr Val Leu Tyr Tyr Tyr Ser Asn Leu Ser Val Pro
65                  70                  75                  80
Ile Gly Arg Phe Gln Asn Arg Val His Leu Met Gly Asp Asn Leu Cys
                85                  90                  95
Asn Asp Gly Ser Leu Leu Leu Gln Asp Val Gln Asp Val Glu
            100                 105                 110
```

<210> 518
<211> 4544
<212> DNA
<213> Homo sapiens

<400> 518

```
ccgagaaggg cttcaggacg cgggaggcgc acttgcttca agtcgcgggc gtgggaacgg    60
ggttgcaaaa cggggccttt ttatccgggc ttgcttccgg cgtcatggct caaagggcct   120
tcccgaatcc ttatgctgat tataacaaat ccctggccga aggctacttt gatgctgccg   180
ggaggctgac tcctgagttc tcacaacgct tgaccaataa gattcgggag cttcttcagc   240
aaatggagag aggcctgaaa tcagcagacc ctcgggatgg caccggttac actggctggg   300
caggtattgc tgtgctttac ttacatcttt atgatgtatt tggggaccct gcctacctac   360
agttagcaca tggctatgta aagcaaagtc tgaactgctt aaccaagcgc tccatcacct   420
tcctttgtgg ggatgcaggc cccctggcag tggccgctgt gctatatcac aagatgaaca   480
atgagaagca ggcagaagat tgcatcacac ggctaattca cctaaataag attgatcctc   540
atgctccaaa tgaaatgctc tatgggcgaa taggctacat ctatgctctt cttttttgtca   600
ataagaactt tggagtggaa aagattcctc aaagccatat tcagcagatt tgtgaaacaa   660
ttttaacctc tggagaaaac ctagctagga agagaaactt cacggcaaag tctccactga   720
tgtatgaatg gtaccaggaa tattatgtag ggctgctca tggcctggct ggaatttatt   780
actacctgat gcagcccagc cttcaagtga gccaagggaa gttacatagt ttggtcaagc   840
ccagtgtaga ctacgtctgc cagctgaaat tcccttctgg caattaccct ccatgtatag   900
gtgataatcg agatctgctt gtccattggt gccatggcgc ccctggggta atctacatgc   960
tcatccaggc ctataaggta ttcagagagg aaaagtatct ctgtgatgcc tatcagtgtg  1020
ctgatgtgat ctggcaatat gggttgctga gaagggata tgggctgtgc acggttctg  1080
cagggaatgc ctatgccttc ctgacactct acaacctcac acaggacatg aagtacctgt  1140
atagggcctg taagtttgct gaatggtgct tagagtatgg agaacatgga tgcagaacac  1200
cagacacccc tttctctctc tttgaaggaa tggctggaac aatatatttc ctggctgacc  1260
tgctagtccc cacaaaagcc aggttccctg catttgaact ctgaaaggat agcatgccac  1320
ctgcaactca ctgcatgacc ctttctgtat attcaaaccc aagctaagtg cttccgttgc  1380
tttccaagga aacaaagagt caaactgtgg acttgatttt gttagctttt ttcagaattt  1440
atctttcatt cagttccctt ccattatcat ttacttttac ttagaagtat ccaaggaagt  1500
cttttaactt taatttccat ttcttcctaa agggagagtg agtgatatgt acagtgtttt  1560
gagattgtat acatatcttc cagaactttg aggaaatctt atttaagttt atgaatataa  1620
ccatctgtta ctgttctaaa aatgtttaaa agaaactcaa tacagataaa gataaatatg  1680
tgactattat tgggtattac acttcacttc tctttaatat ttttcctcca actggagggc  1740
agacaatttt ctgacttgct tttctctagg tggttcattt tgaaagggga cagaaatata  1800
actaaatgct tccaggagaa aaattccaag agttacaatc tggacttggt acctaaatat  1860
cattttttaa attcttgatg cctatttgga ctagaggtaa acatactttc agattggcct  1920
gttttttgtcg gtaaggcata cagccttcag aagccaacat ttttaatcaa aaacttataa  1980
aacatgatga tcattgtgaa aattctgagt tgaaggttag tttaagataa gctaacaata  2040
acagtctgtg ttttctctaa ataatctga gttttttgga actctttatt taaatatgtg  2100
tgttttttcag tattcaaata agatcaggaa gccaattttc tatgtatgaa tatgctttaa  2160
cctaggattt cagtccactc tgactgactt tctaaacttt aacttgggtt tttacagtga  2220
ctatgcatta gtgctgactc tttggtataa gccataaaat attttccttc ctatcaattt  2280
atctgaactt tggtcttttc actaaattgt acagtattct acttctgttt aaaaagggga  2340
gatgagaaag ggaatactat ctaaccaata acttgaacaa aaacactaaa ctaagcattt  2400
aatagaaatg cttttttattg aggaggtatt atccagagtt catgcttaga acaaatgcat  2460
ctttgcgtat cctagactta acaattcatc agtttctgag accacagaat caggttttcc  2520
gtagtagata aagactctct ggtgcttcaa attctgttca agtgttttga ctcatcagct  2580
tctactcttt ctattactgc ctttgcctgg cttgttttgt ctctttgcaa ctgattttgc  2640
aaaaaaaaat tgtagcttta aaataacagg gtctaagtat tttaaatgtg cctatttcac  2700
```

```
agctctcttg gtcacaaaaa catgctattt ttattggaac ttcaaaccaa atccccactg   2760
agtgtgtact ggttcctgca ggtagcagtc tcctattatc tcctgtttag caccaaaaga   2820
gctaatatta ttggaaactg acctttaaa ggccactggc agtaggattt aaaaagcagc    2880
ccactgctca gtttccagga tcagcttcct ccttctgtca cttgtgtaag ttggcactac   2940
cttgtgcctc tcagattgct gaagtgctgc tggtaagcat gtgcatgctc tgcctttctt   3000
gtgaaagttt tcaatcagcg atatcagcac ttacagtaag aagtaaaagt agtgcacagc   3060
aaagctaatt tgcctttgcc tggggtgttc agcttgaaag aataaagctc atttggttta   3120
gttaaatgtc ttactctact gtgcctatgc ttttagctgc gttactaagc aagggaaaaa   3180
taacagtttc tctgagccag agaagacttg atcacagttc tccaagcatc gtgatagcaa   3240
tgcttaaccc caggaagatt tcaaggcagg gagaagaaca tttcaaataa gattcttgtt   3300
aacccattta tgcctagtgt tccattattg gaatgctaag cttgtgggag tcatttacat   3360
cctactgctc aaagtcattg ccaaggtctg attttttcaca caaaaaattg caaccccag    3420
cataaatggg ttagctactg tcatcagtta gcaaattcat ccacacaaac acaattagag   3480
tttggttttt ttttaagctt ttcaaaactt actaaactgg cacaatttta tatgtatgct   3540
atttgttgta tttatgctta agagcaaaaa agttttgatg ggattttaaa ttcagcaaag   3600
cctacaacgc tgagacaatc ccctaacaac atggtagtaa ctaaagaaac ttttatacta   3660
ggcttcttag ttttaaaagg aagtggcatc attgtttcag ttctagtttg tatttttctc   3720
tcagatattt ttcttctttta aaaatctttc ccagaagttg gttcctagaa aactcaatac   3780
catcatctct tatctctata cagggactag gtaataaaac cttcaaaggt tgccaaaggt   3840
catcaagcag tgttcattta tcctgtcaca tgtttctgtt tctatagtaa tttagaaatt   3900
gcaaatagtt aacttttcat catgtaaaaa gttaacatta tcctatttcc atagatacca   3960
tggacggcgg tgtggcctga gttgtcagtc tttaatcctg agtcatgtgg ctctcttttc   4020
atctttgatg tcagttccaa ttatttggca tcaaaaacct tcatggtagg tagagttta    4080
ggtaaaagtg gatctagggt tactttcttt attaacattt cctaaataac tgaattgaga   4140
gacatactct gctactatgt cctcaggtta attttgtct gatcttacga tgccctgcct    4200
tttactagct actttagaaa tagaaaatgt gaagagtgac tatttacatg tatactcctt   4260
tggctgctag aactcatctg tagtcctta ttatttacac tgaattccaa tttcattct     4320
cttccgctaa gtaagagcac ctcattcctg tgttttctct actattgagc tgtagacgaa   4380
ctgtttctct aattataaag caaactgttt gggatattca gggaaactac cccaatgtta   4440
tgttgtcatt taatgggaaa ggctgggatc atatgtattt ctatgttctg taaagtattt   4500
gacttactag ttctcaataa aattttatta ggactataaa aaaa                     4544
```


<210> 519
<211> 1779
<212> DNA
<213> Mus musculus

<400> 519
```
ggtccggaat tcccgggtcg acccacgcgt ccgctggcct tgggcgcaga ccccggccgg     60
tcccggggct gcctctttaa gggaggggt ggagccgcga gtcaggcgcg aggagctcca    120
gaaatcttga ggccagagcc ccgcacctcg gcgcagccat gagtgcggag gtgaaggtga    180
cagggcagaa ccaagagcag tttctgctcc ttgccaagtc ggctaagggg gcggcactgg    240
ccacactcat ccaccaggtg ctggaggccc ctggtgtcta cgtgtttggg gaactgctgg    300
atatgcctaa tgttagagag ctggcagaaa gcgactttgc ctccaccttc cggctgctca    360
cagtgtttgc ctatgggacc tatgcggact acttagctga agccaggaat ctccccccac    420
tgactgacgc acagaagaat aagcttcgac atctgtcagt tgtcactctg gctgccaaag    480
tcaagtgtat cccatatgca gtgttgctgg aggccttgc ccttcgaaac gtgcgccagc     540
tggaagacct tgtgatcgag gctgtgtatg ctgatgtcct tcgtggctct ctggaccagc    600
gcaatcagcg gctagaggtt gattacagca tcgggcggga catccagcgc caggacctca    660
gtgccatcgc ccagaccctg caagagtggt gcgtgggctg tgaggttgtg ttgtcgggca    720
tcgaagagca ggtcagccgt gccaaccagc acaaggagca gcagctgggc ctgaagcagc    780
agatcgaaag tgaggttgcc aaccttaaga aaaccattaa agttacgaca gcagctgctg    840
ctgcagccac ctcccaggat cctgagcaac acctgacaga gctgagagaa ccagcttctg    900
gcaccaacca gcgccagccc agcaagaaag cctccaaggg caagggactc cgagggagcg    960
ccaagatttg tccaagtcg aactgaaagg acttgtttct tccctgggaa tgtggggtcc    1020
cagctgccta cctgcctacc ccttaggagt cctcagagcc ttcctgtgcc cctggccagc   1080
tgataatgct agttcattac ttttcatctc ctccacccce aagcataagc cacaccctct   1140
gtagggagga ggccagtgca ggtcatgttc tgttggtacc tcttatgtgt tccatgctct   1200
tccccagcac gcttgctctc atcgtttttc cgcactgtgt ctgcccatta ccctgtcat    1260
tgagcaggtt ggcagtccta tggagggtgc tggctcttaa ccacccacac ctaccctgc    1320
```

```
atgcctaatc tgcagttcct cctcctcccc ttgcctagtg ggctgcatct gaaaagccat   1380
ggggaagggg gtctccacct tcattccagc cttagagttc tggagccagt ctgctaccct   1440
gggagtcgct ggacattttc ctcccagaac cccatcacac tacaattgtt tctttcctct   1500
ctcatctcct tgggcctggg gatactgctg cttcagtgac cccagagcct gagaacagct   1560
attttttgaga tgttaagaaa tggttctttg ttgctcatca tcttaggaag cccaatggaa   1620
atcctggaag gatttatatc tcctcctgtg gttctggtgg ggaaggaaat atagattgta   1680
tattaaaaat aaaaaatata tatgaatagg tctatatata ttgacacatg acacagaaat   1740
aaatgtatga gaaatgtatg tacaaaaaaa aaaaaaaaa                            1779
```